nature

www.nature.com/nature

Vol 450 | Issue no. 7173 | 20/27 December 2007

Newsmaker of the Year

Nature is pleased to name Rajendra Pachauri, the Indian engineer and economist, and chair of the Intergovernmental Panel on Climate Change, as our inaugural Newsmaker of the Year.

Science is perhaps less reliant on the concept of 'personality' than any other major realm of human endeavour. The scientific facts will remain the same, whoever first discovered them, or described them, or imagined that they might be so.

At the same time, it is because of who scientists are and how they work together, who they like and who they cannot stand, their beliefs about the world, their stamina and their foibles, that research actually gets done. Science, like history, is forged by individuals — even though both are forged on the back of a past whose inhabitants may have faded into anonymity.

It is through people, too, that science shows its public face. When science becomes news, it does so through human agency. That is why at the end of each year, from now on, *Nature* will single out for recognition a person whose role in science has had a particular impact on the wider world stage — a 'Newsmaker of the Year'.

A newsmaker is not necessarily someone to celebrate. In previous years we might have chosen a figure of obloquy, such as Woo Suk Hwang, the disgraced stem-cell researcher. In future years, it is not beyond the bounds of possibility that a cloned human being, a misguided politician or even a bioterrorist could be selected; anyone might have a significant impact in the news and on science itself, and deserve some sort of singular analysis.

But the contribution of this year's winner to scientific affairs can be celebrated without reservation. Rajendra Pachauri's great strength is in building and organizing institutions in the fields he understands best — engineering and economics as they apply to issues of development. In that area he has enjoyed a success that reflects his calm, yet fiercely driven personality (see page 1150). Over two decades he has built TERI, the Delhi-based energy and resources institute that he runs, into an organization with offices around the world and several hundred staff. And in the past five years, he has chaired the great collaboration that is the Intergovernmental Panel on Climate Change (IPCC).

The concept of an annual newsmaker does not signify an infatuation with star power, however. Discoveries, rather than personalities, remain at the core of scientific research. That is why *Nature*'s sister journal *Nature Methods* is introducing a 'Method of the Year'. The first winner is next-generation ultra-rapid DNA sequencing.

Pachauri's year has already featured his receipt, on behalf of the IPCC, of a share in the Nobel Peace Prize. It concludes with the moderately successful completion of the UN Convention on Climate Change talks in Bali earlier this month (see, page 1136), when nations made some headway in determining the likely shape of an agreement to succeed the Kyoto Protocol, which expires in 2012.

Protecting the vulnerable from the threat of climate change is about changing what we all do, and that requires political action as well as changes in personal behaviour. Burying carbon underground and lighting our bedrooms with the power of the atomic nucleus or the tides are things that need to be arranged by governments, both directly — by making the economic costs of carbon emission fall on the processes that emit it — and indirectly, through basic research and spurs to technology development. The Bali meeting provided just a taste of the testing political discourse ahead. Behind that lies the hard reality of the personal costs of mitigating climate change, which will fall alike on those who bear them — whether willingly or unwillingly.

But collective action has a positive and uplifting side, too. The IPCC is a case in point. Its members have sacrificed time that they would rather have spent on new research to do something for the world at large. Their endless meetings and discussions, their intellectual clashes and warm mutual understandings, have produce an unparallelled catalogue of reliable knowledge — and authoritative assessments of remaining ignorance — on a scientific matter of utmost public concern. To produce something that the hundreds of authors can be proud of, and in which the nations of the world have all, to some extent, invested their trust, is no mean thing. The IPCC's collective efforts span decades. But the person sitting in the chair at its hour of greatest achievement so far is Rajendra Pachauri, and we salute him. ■

A policy of drift

British physics faces an unnecessary squeeze.

In an 11 December announcement of the UK research councils' budgets for the next three years, the UK government's innovation secretary, John Denham, called the settlement "good news" for British science. But the numbers were bad news for the Science and Technology Facilities Council (STFC) which, for historical reasons, funds research in particle physics and astronomy, as well as facilities. The council's plan for implementing the budget takes from the former to pay for high operating costs on the latter — with potentially painful consequences for physics departments in UK universities.

After the government published its comprehensive spending review in October, it became clear that the STFC would not receive the necessary funds to absorb these running costs. Instead, the council is facing a funding shortfall of about £80 million (US$160 million) over the next three years. It plans to deal with this by pulling out of the international Gemini telescope project, stopping preparatory work on the proposed International Linear Collider, and slashing

funds available for research grants by 25%. It will also cut support for space-based scientific instruments by one-third, and reduce support for solar physics and high-energy γ-ray astronomy.

Researchers in most disciplines, in most parts of the world, have to tighten their belts from time to time. But these reductions are more drastic and sudden than any arm of a competently managed research agency should have to bear.

The funding shortfall arises in part because UK subscriptions to CERN, the European Southern Observatory and the European Space Agency are increasing owing to the weakness of the pound against the euro, and the relative growth of British gross domestic product, on which the United Kingdom's contributions are based.

But the main cause of the gap is the rising operating costs of the Diamond synchrotron light source and a second target for the ISIS neutron source, both sited at the Rutherford Appleton Laboratory in Oxfordshire. The popular facilities, which are used by researchers in disciplines ranging from biomedical research to condensed-matter physics, are projected to cost around £60 million to operate over the period of the council's plan.

The problem has been looming for some time. But the Department of Innovation, Universities and Skills (DIUS), which oversees the research councils, was only created in June and has been unable to obtain additional funds to ease the STFC's plight. Keith Mason, the council's chief executive, has publicly attributed this to the fact that government officials are insufficiently convinced of the economic value of physicists' and astronomers' work.

It should not always be necessary for scientists to provide a purely economic justification for fundamental research into the nature of the Universe. But that case can be made: this research creates skills and ideas that feed into a stronger society and a stronger economy.

The withdrawal from the linear collider and from Gemini reflect badly on Britain's readiness to stand by international collaborations, and will disappoint partners who had long held the nation and its research councils in high esteem. Moreover, grants are being cut in fields where Britain has traditionally excelled, even as the STFC proposes new projects for which a strong scientific case has not been made — such as a joint robotic Moon mission with NASA.

"Government officials are insufficiently convinced of the economic value of physicists' and astronomers' work."

Both the DIUS and STFC, which was founded from an amalgamation of two research councils only in April, are young organizations and their inability to secure extra funding may reflect their relative lack of proficiency in the ancient art of Whitehall infighting. But it also seems, from Mason's comments, that senior officials at the Treasury do not consider astronomy or particle physics relevant to that department's policy of backing research that will foster business innovation.

Denham has asked a panel chaired by Bill Wakeham, vice-chancellor of the University of Southampton, to review the likely impact of the proposed changes and report next spring, and the House of Commons innovation committee is launching its own enquiry into how the shortfall came about. These reviews should find out whether it is possible for particle physics and astronomy grantees to be treated fairly inside a research council whose priority will always be the provision of facilities. They should also explore ways of ensuring that disciplines using facilities such as Diamond pay their fair share of operating costs. ■

Welcome *Nature Geoscience*

Not before time, the world's population is focusing its attention on threats arising from humankind's impacts on its planetary habitat. But underlying whatever abilities we have to mitigate the impacts of habitat change lies the essential ability to understand our planet's structure and dynamics at all levels and on all time-scales.

The slow but steady progress in understanding Earth is the momentous achievement of the geosciences since their foundations were laid well over a century ago. *Nature Geoscience* is *Nature*'s newest monthly sister journal, and is intended to capture the best of those sciences and serve all of the interested research communities (see http://www.nature.com/ngeo). As with all Nature research journals, this launch in no way dilutes *Nature*'s own commitment to these disciplines.

Launching in January 2008, the journal can be expected to track important research currents. For example, an expanding branch of climate science is the investigation of past analogues for current change, such as rapid sea-level rise, warm periods, ocean acidification or such crucial links as the coupling between atmospheric carbon dioxide concentrations and global temperatures over geological time-scales. Traditionally rooted in geology, palaeoclimate researchers are now interacting more intensively with modellers and with investigators of modern climate.

Solid-Earth scientists are still exploring the implications of the 2004 discovery of the high-pressure phase of the mineral $MgSiO_3$, the post-perovskite that dominates the lowermost mantle, for our understanding of the structure of this region. This discovery affects our understanding of the temperature in the lower mantle and mantle convection, as well as iron and heat exchange between Earth's core and the mantle. All these, in turn, have implications for our reconstructions of Earth's history, the evolution of its core and the geodynamo.

Like all the Nature research journals, *Nature Geoscience* will also have its informal aspects. The 'Backstory' section, about the hard work that comes before a publishable research paper, will describe, for example, what it takes to reconstruct 15 million years of Arctic ocean circulation, to map the floor of the Arabian Sea or to study an Alaskan glacier. One of these articles will be published on the final page of each printed issue, with additional articles published online each month.

The coming year sees the start of a period of international programmes focusing on our planet, not least the International Year of Planet Earth. The new journal could not arrive at a more auspicious moment. ■

RESEARCH HIGHLIGHTS 2007

As an end of the year round-up, we asked *Nature*'s editors to nominate favourite papers published elsewhere this year. For a pick of favourites from *Nature* itself, see page viii.

MOLECULAR BIOLOGY

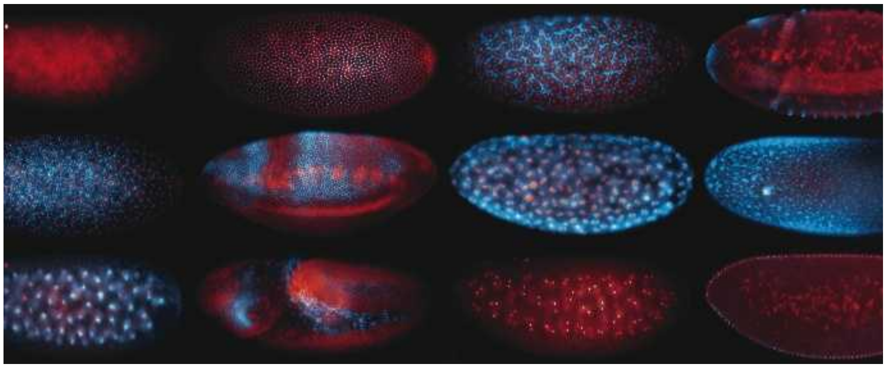
Stem cell success

Cell **131,** 861–872 (2007); *Nature Biotechnol.* doi:10.1038/nbt1374 (2007); *Science* doi:10.1126/science.1151526 (2007)

In 2006, Shinya Yamanaka and Kazutoshi Takahashi of Kyoto University in Japan successfully prepared stem-like cells, which can develop into almost any of the body's cell types, from adult mouse skin cells. The technique involved 'reprogramming' the skin cells by introducing four specific genes using retroviruses, and provided a means of obtaining stem-like cells without using eggs or destroying embryos.

In 2007, Yamanaka and his colleagues went one better, successfully reproducing the experiment with human skin cells. The recipe works because the additional genes function as high-level regulators, turning on other genes and thereby setting off a cascade of intracellular changes.

Later, the group reported that it had eliminated the cancer-causing gene *c-Myc* from the list of introduced genes, a research boon along the path towards patient-matched stem-cell therapies. James Thomson of the University of Wisconsin-Madison and his co-workers also published a *c-Myc*-free method.



H. KRAUSE

RNA on the move

Cell **131,** 174–187 (2007)

Moving bulky proteins around a cell takes a lot of energy. These costs can be cut by moving a single messenger RNA (mRNA) and then producing many copies of its encoded protein where they are needed. Henry Krause of the University of Toronto and his colleagues have revealed that this happens to a much greater extent than expected in the early fruitfly (*Drosophila melanogaster*) embryo. The researchers followed the movements of mRNAs during the first 4.5 hours of embryonic development. They found that 71% of the mRNAs in their sample were moved to specific places.

The monitored mRNAs were targeted to about 35 different regions, and each type was always sent to the same places. Furthermore, the distribution of proteins closely matched patterns of mRNA localization. These results suggest that mRNA localization could regulate many cellular functions.

CANCER

Aid to resistance

Science doi:10.1126/science.1141478 (2007)

The drug gefitinib is at first an effective treatment for certain types of lung cancer, but most cancers become resistant to it. New research provides insight into one mechanism by which this happens.

Gefitinib inhibits an enzyme known as EGFR kinase, and around half of the resistance cases are due to mutations in this target. Among the rest, Pasi Jänne of the Dana-Farber Cancer Institute in Boston, Massachusetts, and his colleagues found that some resistant cells showed increased expression of a tyrosine-kinase receptor named MET. This activates the same signalling pathway that EGFR kinase triggers — a pathway that mediates cell survival.

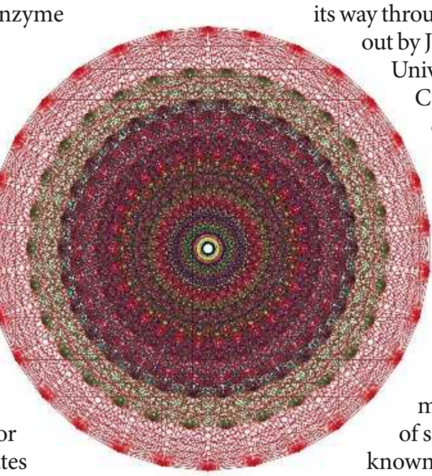
The researchers found that the MET gene had multiplied rather than mutated, and that inhibiting MET restored the cells' sensitivity to gefitinib.

MATHEMATICS

The 248th dimension

www.liegroups.org

It took 77 hours for a supercomputer to crunch its way through the calculations set out by Jeffrey Adams, of the University of Maryland in College Park, and his co-workers. The result was a matrix with 205 billion entries that describes the 248 axes of symmetry and 57-dimensional surface of a mathematical structure called E_8 (represented left). E_8 is the largest and most complicated object of smooth symmetry, known to mathematicians as an 'exceptional Lie group'. The team's main challenge was to work out how to condense the sums required so as to make them small enough for a supercomputer to handle, a feat achieved by invoking the 'Chinese remainder theorem'.



P. MCMULLEN/AM. INST. MATH.

MAGNETIC IMAGING

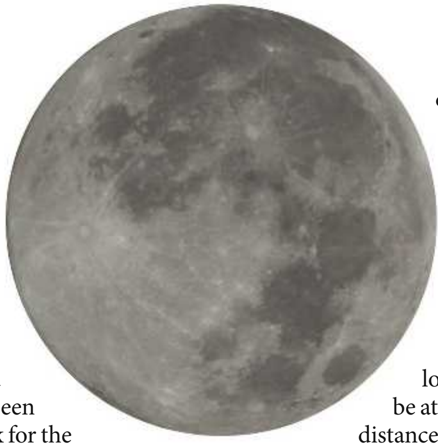
Getting to the point

Nature Nanotechnol. **2,** 301–306 (2007)

The resolution achieved by magnetic resonance imaging (MRI) has improved 60,000-fold, thanks to developments that couple MRI with atomic force microscopy.

In magnetic atomic force microscopy, a sample, in the form of a thin film, is placed on an ultrasensitive cantilever that is suspended above a conical magnetic tip. The tip's strong magnetic field means that changes in the spin of atomic nuclei in the sample change the way the cantilever oscillates.

John Mamin, at IBM's Almaden Research Center in San Jose, California, and his colleagues have developed highly magnetized tips with extremely sharp points, which generate large magnetic-field gradients. This allows them to achieve a resolution of 90 nanometres, making the system vastly more sensitive.



NASA

PLANETARY SCIENCE

The whole of the Moon

Earth Planet. Sci. Lett. **262,** 438–449 (2007)

The consensus that Earth's satellite was created when a planet sometimes called Theia hit the early Earth has been tormented by a paradox for the past 30 years. Although computer simulations of the impact suggest that most of the Moon's mass came from Theia, its oxygen-isotope composition is very similar to that of Earth, suggesting that in fact it broke off from our home planet.

Kaveh Pahlevan and David Stevenson of the California Institute of Technology in Pasadena believe that their model resolves this contradiction. They propose that the massive impact some 4.5 billion years ago vaporized most of both Earth and Theia. The disk of magma that eventually became the Moon remained linked to Earth by a wash of silicate vapour. Within this vapour, turbulent mixing evened out the isotopic differences between Earth and Theia before the Moon condensed out, the researchers suggest.

QUANTUM CRYPTOGRAPHY

Canary communication

Nature Phys. doi:10.1038/nphys629 (2007)

Physicists demonstrated the ability to distribute a quantum 'key' over a record 144 kilometres, the distance between two telescopes in the Canary Islands.

Anton Zeilinger and Rupert Ursin of the University of Vienna in Austria and their colleagues used a laser to produce two photons 'entangled' in such a way that a measurement of one influences measurements of the other. Quantum cryptography makes use of this effect to distribute cryptographic keys no eavesdropper could intercept without being detected.

While one photon was measured on the island of La Palma, the second was picked up by a telescope on the island of Tenerife. The team hopes that this advance will one day lead to satellite-based quantum communication.

PARTICLE PHYSICS

Hard-core calculations

Phys. Rev. Lett. **99,** 022001 (2007)

Why do protons and neutrons stick close enough to form an atomic nucleus but never merge into a formless soup of their constituent quarks?

In July, a group of Japanese physicists answered this question by publishing first-principles calculations of the particles' internal dynamics.

The force between nuclear particles has long been known to be attractive at moderate distances but repulsive at very close quarters, yet quantitative explanations of this in terms of the behaviour of the particles' constituent quarks have proved elusive.

Now Noriyoshi Ishii of the Universities of Tsukuba and Tokyo and his colleagues have used sophisticated algorithms and massively powerful computers to derive the familiar form of the nuclear force from the quantum field theory of quark interactions, called quantum chromodynamics. The achievement is both a computational *tour de force* and a triumph for theory.

ECOLOGY

Hot threesome

Science **315,** 513–515 (2007)

A virus within a fungus is vital if that fungus and the grass it lives in are to survive scalding soil temperatures, according to work by Marilyn Roossinck of the Samuel Roberts Noble Foundation in Ardmore, Oklahoma, and her colleagues. They studied *Dichanthelium lanuginosum*, which grows in soils heated by geothermal activity in America's Yellowstone National Park (pictured below). Plants that lack the fungus *Curvularia protuberata* cannot survive in soils above 38 °C.

The researchers detected a virus inside the fungus. When plants were infected by a fungus that had been completely 'cured' of the virus, they shrivelled and died at high temperatures. Restoring the virus restored the system's heat tolerance. The virus–fungus combination also protected some tomato plants against high soil temperatures.

CELL BIOLOGY

Barbed wires

Cell **128,** 901–913 (2007)

Inside a cell, networks of filaments made from a protein called actin push and pull on membranes to generate shape and movement. Jack Taunton of the University of California, San Francisco, and his colleagues found that a protein required for creating new filaments is also involved in attaching them to membranes.

This protein, known as N-WASP, captures the fast-growing, 'barbed' end of actin filaments and tacks them onto membranes. The authors describe how mutations in two parts of the protein called WH2 domains stop this from happening and cause actin filaments to separate from membranes *in vitro*.

The team also found that intact versions of these domains were needed for cancer cells to form specialized protrusive structures called podosomes. Taunton says these podosomes may be necessary for cancers to spread around the body.

OCEANOGRAPHY

Atlantic abatement

Science 317, 935–938 (2007)

Fears that global warming might be shutting down the circulation of the North Atlantic were allayed by the initial results from an array of instruments that transects the Atlantic between the Canary Islands and Florida.

In the Atlantic deep, cold water flows south while warm, near-surface water flows north, and this 'overturning circulation' is of profound climatic importance. Stuart Cunningham of Britain's National

T. CRADDOCK/SPL

Oceanography Centre in Southampton, Jochem Marotzke of the Max Planck Institute for Meteorology in Hamburg, Germany, and their colleagues accounted for an earlier report of a 30% decrease in the rate of this overturning, which was published in *Nature* in 2005 (H. L. Bryden *et al. Nature* **438,** 655–657; 2007). Having studied variations in circulation over time, the researchers write that the earlier conclusion was based on a measurement taken during a natural lull, a possibility that was acknowledged in the original study.

ASTRONOMY

Sources in the sky

Science **318,** 938–943 (2007)

A collaboration involving more than 440 researchers associated with the Pierre Auger Observatory in Mendoza, Argentina, has traced ultra-high-energy cosmic rays to their apparent source for the first time.

Unlike lower-energy cosmic rays, these ultra-high-energy particles are only marginally deflected by magnetic fields they encounter on the way from their sources to Earth. Using 1,600 ground-based instruments spread across 3,000 square kilometres of Argentina's Pampa Amarilla plain, the Auger team was able to gather enough information on the extremely rare rays to say where in the sky they were coming from. The sources coincide with previously mapped active galactic nuclei (see illustration below), thought to be powered by supermassive black holes.

A. SIMONNET, SONOMA STATE UNIV.

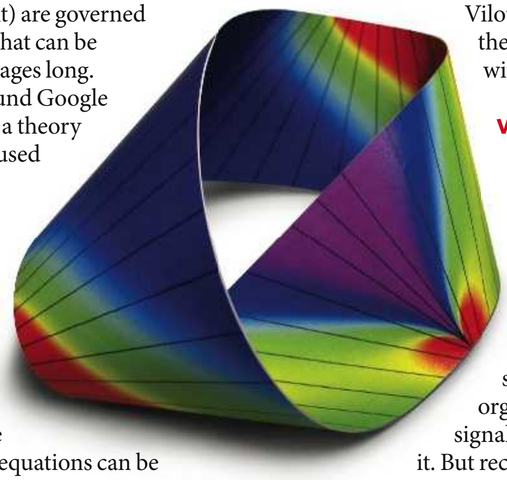
MATHEMATICS

Mind over Möbius

Nature Mater. doi:10.1038/nmat1929 (2007)

Gert van der Heijden and Eugene Starostin of University College London calculated the 'relaxed' shape of an elastic Möbius strip.

The Möbius strip is the one-sided surface with no ends. The final shape and the potential energy stored in its bends (pictured right) are governed by equations that can be hundreds of pages long.



E. STAROSTIN & G. VAN DER HEIJDEN

Poking around Google led the duo to a theory that could be used to drastically simplify the problem by changing the coordinate system in which the calculations are made. The reformulated equations can be used to study a variety of problems, such as the crumpling of paper and the deformation of nanoribbons or twisted biofilaments.

PHYSICS

Gravity's boost

Phys. Rev. Lett. **98,** 021101 (2007)

Dan Kapner and his colleagues at the University of Washington in Seattle measured the force of gravity between two masses only 55 micrometres apart, roughly the width of a human hair. The force rose as the distances between the masses shrank, in accordance with theory.

In quantum field theory, forces are carried by particles. If the force had fallen away faster than theory had predicted, this could have indicated that the particles associated with gravity are relatively large, and that, in turn, might have explained why cosmologists have found much less cosmological 'dark energy' than quantum theory predicts. These authors' findings eliminate that possibility.

CHEMISTRY

A simple solution

Science **317,** 1189–1192 (2007)

Our understanding of how molecules will react is usually gleaned from experiments carried out with organic solvents available in standard chemistry labs. For natural products that come from water-living organisms, it might help to think about things differently.

Timothy Jamison and Ivan Vilotijevic at the Massachusetts Institute of Technology in Cambridge have synthesized the core piece of a famously hard-to-make marine molecule by working in neutral water. The molecule, which causes the toxicity associated with 'red tide' algal blooms, has a ladder-like arrangement of rings. It was thought that this structure might assemble through a cascade of reactions, but not until Jamison and Vilotijevic tried it in water did the molecule zip together with ease.

VISION

A scaffold in new light

Cell **131,** 80–92 (2007)

The fruitfly protein INAD had long been considered to be a scaffolding protein, organizing important visual signalling proteins that attach to it. But recent research suggests that INAD directly regulates visual perception.

Rama Ranganathan, of the University of Texas Southwestern Medical Center in Dallas, and his colleagues show that, in response to light, one of five structural 'PDZ' domains of INAD transiently switches from a reduced to an oxidized state, distorting INAD's ability to bind to other molecules. This seems crucial to visually mediated reflex behaviours and for terminating visual responses.

Many scaffolding proteins contain PDZ domains, which could undergo similar conformational changes to that of INAD. Thus, rather than support components, these might serve as control centres for other signalling molecules.

CANCER BIOLOGY

Shock strategy

Cell **130,** 1005–1018 (2007)

A protein called heat-shock factor 1 (HSF1) that helps cells to handle stress seems to have a sinister alter ego. The latest findings suggest that the protein aids tumour growth.

Susan Lindquist of the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts, and her colleagues studied mice that lack the *Hsf1* gene. These mice developed fewer tumours when they had cancer-linked mutations or when they were exposed to carcinogens than did normal mice. HSF1 also aided human tumour-cell growth in culture.

The results have mixed therapeutic

Pikes and planning

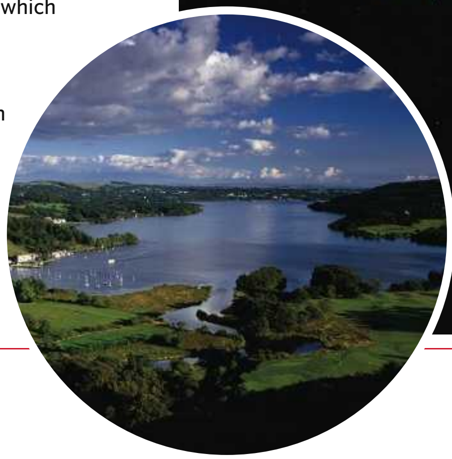
Proc. Natl Acad. Sci. USA **104,** 15799–15804 (2007)

Fisheries management uses population models that assume fishing does not cause rapid evolutionary change. Nils Stenseth of the University of Oslo and his colleagues found evidence against this premise.

The team analysed data on half a century of pike (*Esox lucius*, pictured) catches from Lake Windermere in England's Lake District. The records included each pike's age, which can be counted from rings that form in bones around the gills.

Predators normally prefer the smallest fish, whereas humans prize the largest. So fishermen apply an opposing selective pressure from the rest of nature. During periods when fishing was heavy, it became the dominant pressure, and pike tended to be smaller and to devote less energy to reproduction. The findings suggest that fisheries managers should include the evolutionary effects of fishing in their estimates.

LUTRA/NHPA



A. COOPER/CORBIS

implications. They suggest that HSF1-inhibiting drugs could combat cancer. But HSF1-stimulating therapies are being explored as treatments for disorders including Parkinson's disease. Researchers will need to explore how each approach affects the problem targeted by the other.

CHEMISTRY

Getting the right version

Science **317,** 496–499 (2007)

Mirror-image versions of some molecules — called chiral — are not structurally identical, and ensuring that only one version is produced in a chemical reaction is often a problem. One way is to use a metal catalyst attached to a ligand molecule that is also chiral.

Ligands are normally attached to metal catalysts by strong covalent bonds, but Dean Toste and his colleagues at the University of California, Berkeley, have successfully used negatively charged ligands that are only weakly ionically attracted to the positively charged metal. This means that a small group of charged ligands could make a vast number of existing metal-catalysed reactions chirally selective.

Toste reports three different gold-catalysed transformations that gave more than 90% yield of one chiral product.

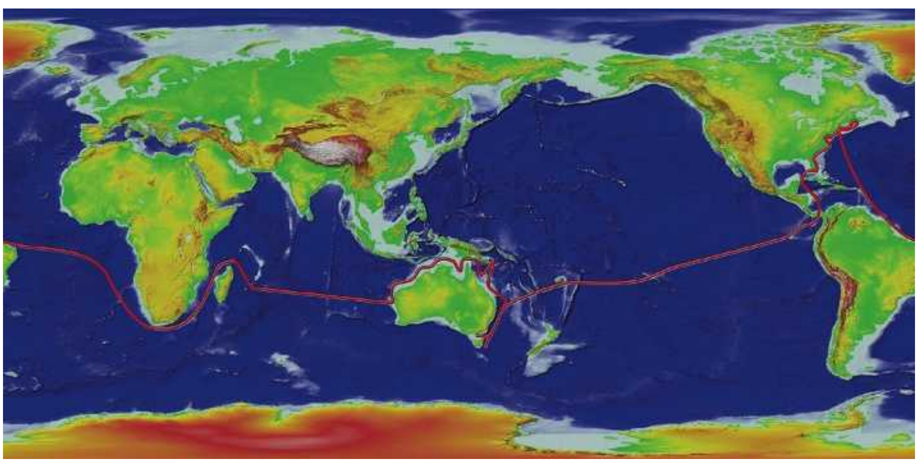
METAGENOMICS

Sorcerer's survey

PLoS Biol. **5,** e16 (2007)

Shibu Yooseph of the J. Craig Venter Institute in Rockville, Maryland, and his colleagues have revealed more of the oceans' genetic diversity than has ever previously been seen.

Sea-water samples collected by Venter's yacht, *Sorcerer II*, on a trip around the world (route pictured below left) contained gene fragments predicted to represent more than six million proteins. In March, when the work was published, that was enough to almost double the number of proteins listed in online databases. The data set included new members of nearly all known protein families from bacteria and archaea, and a large number of viral genes, suggesting viral diversity has to date been undersampled.



J. CRAIG VENTER INSTITUTE

MOLECULAR NEUROBIOLOGY

Active resilience

Cell **131,** 391–404 (2007)

A molecule known as BDNF may provide a pivotal distinction between people who succumb to conditions such as depression or post-traumatic stress disorder and the majority who do not.

Eric Nestler of the University of Texas Southwestern Medical Center in Dallas and his team had previously identified mice that avoided social contact with cage-mates after repeated 'social defeat' — brought about by forced encounters with more aggressive mice.

The researchers found that resilience to stress in this social-defeat model is an active process in which stress-induced firing in part of the brain's neuronal circuitry associated with reward and drug addiction is suppressed. This lowers levels of BDNF (brain-derived neurotrophic factor), and mice with a natural variation in their BDNF gene were resilient to stress. The researchers also found high levels of BDNF in a region associated with reward in post-mortem brains of people with a history of depression.

THIS YEAR IN NATURE

***Nature*'s manuscript editors made a selection of 'favourites' from the papers we published in 2007.**

STEM CELLS

Now in primates

Producing primate embryonic stem cells by somatic cell nuclear transfer, J A Byrne *et al. Nature* **450**, 497–502 (22 November).

This important Article brought somatic cell nuclear transfer to primates for the first time. A group led by Shoukhrat Mitalipov generated two embryonic stem cell lines from 314 oocytes taken from 14 rhesus monkeys. The hope is that this approach could work in humans for generating patient-derived embryonic stem cells. [doi: 10.1038/nature06357]

GENOMICS

Disease risks

Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls, The Wellcome Trust Case Control Consortium *Nature* **447**, 661–678 (7 June).

An extensive genome-wide association study in the British population for bipolar disorder, coronary artery disease, Crohn's disease, hypertension, rheumatoid arthritis, and type I and type II diabetes. The analysis confirms previously identified loci and provides strong evidence for many novel disease susceptibility loci. [doi: 10.1038/nature05911]

QUANTUM PHYSICS

Travelling light

Coherent control of optical information with matter wave dynamics, N S Ginsberg, SR Garner & L V Hau *Nature* **445**, 623–626 (8 February).

This development in quantum control may find application in quantum information processing. The cover illustration represents a unique experiment: a light pulse stopped and extinguished in one 'box' is revived from a completely different box in a separate location and sent back on its way. The 'boxes' in the actual experiment were Bose–Einstein condensates, 160 μm apart. Information was transferred by converting the optical pulse into a travelling matter wave more amenable to manipulation than light. [doi: 10.1038/nature05493]

EXTRASOLAR PLANETS

Caught in transit

A spectrum of an extrasolar planet, L J Richardson, D Deming, K Horning, S Seager & J Harrington *Nature* **445**, 892–895 (22 February).

Fourteen of the 200 plus known extrasolar planets exhibit transits in front of their parent stars as seen from Earth. In theory, subtracting the spectrum seen with the planet 'behind' its star from that seen when it is 'in front', should leave the actual spectrum of an extrasolar planet. This year practice caught up with theory, and an infrared spectrum was obtained for the transiting 'hot Jupiter' HD 209458b. [doi: 10.1038/nature05636]

LINGUISTICS

Lost words

Quantifying the evolutionary dynamics of language, E Lieberman, J-B Michel, J Jackson, T Tang & M A Nowak *Nature* **449**, 713–716 (11 October).

An unusual paper for *Nature* perhaps. A calculation of the rate at which a language grows more regular, based on 1,200 years of English usage. The trend follows a simple rule: a verb's half-life scales as the square root of its frequency. Irregular verbs that are 100 times as rare regularize 10 times faster. Exceptional forms are gradually lost. Next to go, and next to tumble in the cover 'hour-glass', is the word 'wed'. [doi: 10.1038/nature06137]

MATERIALS

Good on the flat

The structure of suspended graphene sheets, J C Meyer *et al. Nature* **446**, 60–63 (1 March 2007).

Graphene — a one-atom-thick layered form of graphite — is a hot topic in materials science and also in condensed matter physics, where it is a popular model system for investigation. This experiment suspended individual graphene sheets over a microscale scaffold so that transmission electron microscopy and diffraction could be used in structure determination. The 'wavy' structure that emerged is a step towards an answer to the question of why such a 'two-dimensional' structure can exist at all. [doi: 10.1038/nature05545]

CANCER GENOMICS

A feat of sequencing

Patterns of somatic mutation in human cancer genomes, C Greenman *et al. Nature* **446**, 153–158 (8 March 2007)

A series of outstanding papers on large-scale cancer genomics have appeared this year. In this example a sample of 518 kinases associated with more than 200 different cancers were chosen for a major sequencing effort. More than 1,000 previously unknown mutations linked to tumour formation — some as 'passengers' that don't contribute to cancer formation, but over 100 of them as 'driver' mutations that do contribute to disease development. [doi: 10.1038/nature05610]

MECHANOCHEMISTRY

At a push

Biasing reaction pathways with mechanical force, C R Hickenboth *et al. Nature* **446**, 423–427 (22 March).

Mechanical force joins heat, light, pressure and electrical potential as a means of kick-starting a chemical reaction by pushing reactants over an energy barrier. In specially designed polymers exposed to ultrasound, rearrangement reactions are accelerated and reaction pathways biased to yield products not obtainable from heat- or light-induced reactions. This work offers new ways of controlling chemical reactions and may lead to mechanically adaptable polymers that can signal impending damage, undergo structure modification to slow the rate of damage, or even self-repair. [doi: 10.1038/nature05681]

NEUROSCIENCE

Leading light

Multimodal fast optical interrogation of neural circuitry, F Zhang *et al. Nature* **446**, 633–639 (5 April).

A powerful new technique for controlling neural circuits that is likely to have an impact in all fields of neurobiology. It uses a light-activated chloride pump derived from *Archea* to label neurons, which can then be inhibited by exposure to light. This complements an existing tool

for activating neurons via a photoactivatable algal channel. As the channels are sensitive to different wavelengths of light, fast bidirectional control over neural activity can be achieved in a single neural circuit. [doi: 10.1038/nature05744]

ATTOPHYSICS

Solid progress

Attosecond spectroscopy in condensed matter, A L Cavalieri *et al. Nature* **449**, 1029–1032 (25 October).

Electron dynamics on the attosecond timescale (an attosecond is a billionth of a billionth of a second) used to be directly measured only in atomic gases. This paper reported the first attosecond spectroscopic measurements in a solid. The ability to time electrons moving in solids over merely a few interatomic distances makes it possible to probe the solid-state electronic processes occurring at the theoretical ultimate speed limit for electronics. [doi: 10.1038/nature06229]

APPLIED PHYSICS

Weighing in

Weighing of biomolecules, single cells and single nanoparticles in fluid, T P Burg *et al. Nature* **446**, 1066–1069 (26 April).

Tiny particles — molecules included — can be weighed with remarkably high resolution using nanoscale mechanical resonators. But not in the presence of fluids, which dampen the vibrations that make the system work. This rules out practical applications such as medical diagnostics or environmental monitoring. But here is an ingenious way around the problem: 'hide' the fluid inside the resonator. A vacuum-packaged resonator holds the solution with particles of interest in microfluidic channels, and weighs single nanoparticles and bacteria at subfemtogram resolution. [doi: 10.1038/nature05741]

MOLECULAR BIOLOGY

Polymerase structures

Structural basis for transcription elongation by bacterial RNA polymerase, D G Vassylyev *et al. Nature* **448**, 157–162 ; Structural basis for substrate loading in bacterial RNA polymerase, D G Vassylyev *et al. Nature* **448**, 163–168 (12 July).

Two papers in the classic mould of structural biology. The first establishes the crystal structure of bacterial RNA polymerase bound to the DNA template and RNA product, revealing a detailed view of the transcription elongation complex. And the second determines the structures of bacterial RNA polymerase elongation complexes bound to NTP substrate analogues with an antibiotic, revealing the mechanism of substrate loading and antibiotic inhibition. [doi: 10.1038/nature05932; doi: 10.1038/nature05931]

MOLECULAR BIOLOGY

A no-nonsense drug

PTC124 targets genetic disorders caused by nonsense mutations, E M Welch *et al. Nature* **447**, 87–91 (3 May)

Many inherited diseases result from premature termination during translation of an mRNA into protein. Muscular dystrophy is one such disease. This work shows that a small molecule, PTC124, enables the translation machinery to bypass sites that cause premature termination, but still terminate normally at the end of the mRNA. PTC124 can restore normal translation of the gene that is mutated in muscular dystrophy, and it also restores muscle function in a mouse model of the disease. This drug offers hope that a wide variety of diseases with similar translation defects might be amenable to treatment that will restore protein function. [doi: 10.1038/nature05756]

CELL BIOLOGY

Architectural prize

Determining the architectures of macromolecular assemblies, F Alber *et al. Nature* **450**, 683–694; The molecular architecture of the nuclear pore complex, F Alber *et al. Nature* **450**, 695–701 (29 November).

The proteomics based technology described in the first of these two papers will allow cell biologists to look at the detailed structure of all manner of macrocellular machines. The second paper shows the power of the technique by tackling the architecture of nuclear pore complexes, the macromolecular assemblies that selectively transport cargo across the nuclear envelope. [doi: 10.1038/nature06404; doi: 10.1038/nature06405]

THE SOLAR SYSTEM

A touch of Venus

Venus Express package, *Nature* **450**, 629–662 (29 November)

The 29 November issue included eight research papers presenting results from ESA's Venus Express mission — which has been in orbit since April 2006. Subjects covered included the atmosphere, polar features, interactions with the solar wind and the controversial matter of venusian lightning. [doi: 10.1038/nature06432]

STRUCTURAL BIOLOGY

Structural work

Crystal structure of the sodium–potassium pump, J P Morth *et al. Nature* **450**, 1043–1049 (13 December).

One of a package of three papers on the structure of P-type ATPases, this Article reports the long-awaited crystal structure of the Na^+,K^+-pump at a resolution of 3.5 Å. P-type ATPases are cation pumps of fundamental importance for all eukaryotes and many prokaryotes. [doi: 10.1038/nature06419]

BIOFUELS

Steady on the alcohol

Production of dimethylfuran for liquid fuels from biomass-derived carbohydrates, Y Román-Leshkov, C J Barrett, Z Y Liu & J A Dumesic *Nature* **447**, 982–985 (21 June).

Ethanol has its limitions as a biofuel: it is highly volatile, absorbs water and has a low energy density. A potentially better liquid biofuel on those three counts is 2,5-dimethylfuran (DMF). This two-step catalytic process, still in the development stage, can produce DMF from fructose, which can be made either directly from biomass or from glucose. [doi: 10.1038/nature05923]

CLIMATE

The human factor

Detection of human influence on twentieth-century precipitation trends, X Zhang *et al. Nature* **448**, 461–465 (26 July)

Climate models suggested that human activity has caused changes in precipitation on a global scale, but no evidence had been found to support the prediction. This paper produced that evidence. A comparison of observed changes in precipitation over land during the twentieth century with climate simulations points to a detectable influence on the latitudinal patterns of precipitation. Anthropogenic factors contributed to moistening in Northern Hemisphere mid-latitudes, but elsewhere, for instance in the Northern Hemisphere tropics, the effect was drying. [doi: 10.1038/nature06025]

NEWS 2007: THE YEAR IN WHICH ...

Words: Daniel Cressey

M. STOJKOVIC/SPL

Pluripotent stem cells created from human skin

Shinya Yamanaka of Kyoto University in Japan and his colleagues and James Thomson at the University of Wisconsin, Madison, and his co-workers separately managed to create pluripotent stem cells from human skin. Reported on the same day in November, the cells can differentiate into any tissue type, and raise hopes that regenerative medicine will be possible from cells other than human embryonic stem cells.

Nobel Peace Prize goes to climate change

In the year it released three reports and an influential synthesis document on climate change, only winning a Nobel prize could have given the Intergovernmental Panel on Climate Change (IPCC) a higher profile. And that is exactly what happened in October. The IPCC shared its prize with climate activist and former US-presidential hopeful Al Gore.

World goes crazy for a baby polar bear

After being abandoned by his mother in 2006, the Berlin Zoo polar-bear cub Knut hit the headlines in March when animal-rights extremists demanded he be put down. Being raised by humans is worse than death for a wild animal, they claimed. Knut survived to celebrate his first birthday this December, having generated a year's-worth of headlines from Norway to Australia.

S GALLUP/GETTY

Earthquake triggers radiation fears

Japan was shaken twice in July; first by a magnitude 6.8 earthquake and then by the discovery that the quake had caused a radiation leak at the Kashiwazaki-Kariwa nuclear reactor. Although the plant's owner insisted the that leak had not led to dangerous radiation levels and posed no threat to people, the incident still caused concern over the safety of Japan's 55 operating nuclear reactors.

China's head of drug regulation is executed

Zheng Xiaoyu, former head of China's food and drug regulatory agency, was executed on 10 July. Xiaoyu had been convicted of corruption over allegations he took 6.49 million yuan (US$85,000) for pushing questionable drugs through the approval process. The execution followed a series of scandals over the approval of counterfeit medicines by the State Food and Drug Administration that led to the deaths of dozens of Chinese citizens.

Avandia drug troubles

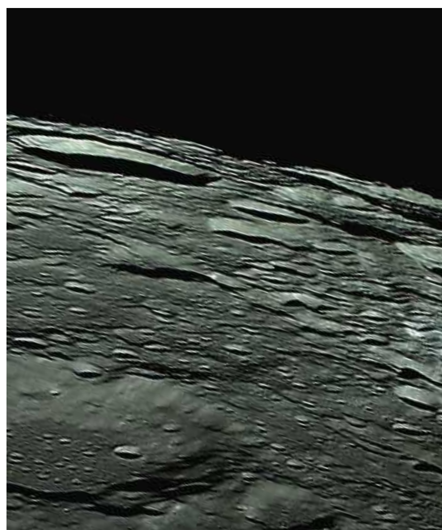
A meta-analysis of data on the diabetes drug Avandia (rosiglitazone) triggered concerns about the drug's safety in May. As well as linking the drug to an elevated risk of heart problems, the study also sparked debate about the use of meta-analyses in medicine. The US Food and Drug Administration later demanded that new warnings be attached to boxes of Avandia and changed the prescribing information for the drug, detailing the new risks.

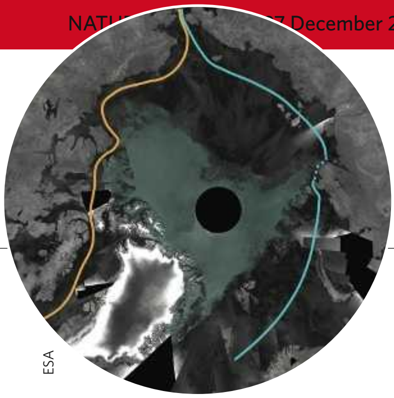
First whole human genome decoded

James Watson, co-discoverer of the structure of DNA, and genomics pioneer Craig Venter announced that their full genomes had been sequenced. The achievements were the first in an anticipated wave of personal-genome sequencing and crucial steps towards personalized medicines tailored to an individual's genetic makeup. Watson's genome, announced in June, was analysed by Connecticut company 454 Life Sciences' new rapid-sequencing technique. Venter's, published in September, was the first fully sequenced diploid genome — detailing DNA inherited from both parents — and revealed that human genetic variation is greater than previously thought.

In November, Google-backed Californian biotech firm 23andMe launched a $1,000 personal genome service; the same month that Icelandic deCODE genetics offered DNA testing for disease-linked genes for the same price. And 2007 saw a splurge of research papers from genome-wide disease-association studies, including diabetes and cancer.

J. SULLIVAN/GETTY





ESA

2007 IN REVIEW
Revisit our best stories in our year-end special, from 21 December.
www.nature.com/news/specials/index.html

Arctic ice shrinks to record low

Arctic sea ice retreated to an all-time low, covering just over 4 million square kilometres in September. This year's melt also opened the long-coveted Northwest Passage shipping route, which connects the Atlantic and Pacific Oceans.

Chikyu drilling ship sets sail

Japan's gigantic research ship Chikyu set off on its first scientific expedition. The 210-metre, US$526-million drilling vessel completed its first mission in November. Its long-term goal is to obtain six kilometres of rock samples from the Nankai trough earthquake zone off the coast of Japan.

Biolab security blunders

The United Kingdom's farming community was rocked by an outbreak of foot-and-mouth disease in August, caused by a biosecurity lapse at an animal lab in Surrey. Subsequent inspections pointed to a leaky pipe on the site, the responsibility for which seemed to fall between the government-owned facility and a privately run lab that share the site.

In June, Texas A&M University was ordered to stop research into bioweapons after an investigation by a pressure group revealed that researchers there had been exposed to dangerous bacterial agents and toxins in its labs.

Geysers disappear … and return

Some of the 41 water displays in Russia's Kamchatka Valley of the Geysers re-emerged in September, three months after being buried in a landslide.

V. SUVOROV/AFP/GETTY

Libyan medical team freed

After eight years in jail, medical workers sentenced to death in Libya for deliberately infecting children with HIV were freed by a consortium including the eleventh-hour, scene-stealing involvement of French President Nikolas Sarkozy and his wife.

The Palestinian-born doctor and five Bulgarian nurses were incarcerated after an outbreak of HIV in Al-Fateh Hospital in Benghazi in 1998, despite a lack of scientific evidence. The case met with international condemnation and outrage by scientists.

The medical workers' sentences were commuted to life imprisonment and they were flown to Bulgaria, where they were pardoned on the airport tarmac in Sofia in July.

B. KATSAROVA/AFP/GETTY

China destroys its own satellite

China sparked international star wars fears after testing a space weapon on one of its own satellites. According to sources in the United States, a Chinese vehicle deliberately smashed into the redundant Feng Yun 1C weather satellite in January. Debris from the remains of the satellite and its assassin may endanger future satellites in similar orbits, experts warned.

Storms rage at the National Hurricane Center

Bill Proenza was appointed director of the National Hurricane Center in Miami in January and soon became involved in a political spat. He ostracized his employers at the National Oceanic and Atmospheric Administration and many of his own staff, who demanded he go. In July, he was placed on leave and he has since left the centre.

Asian nations reach for the Moon

Japan and China joined the growing band of countries intent on grabbing a piece of Moon-related action. Both nations launched probes that returned photographs to Earth, although China was then forced to deny that it had faked its shots.

Private companies also looked to the Moon after Google announced a US$30 million lunar X-Prize in September for whoever could land a robot there and complete a series of tasks before 2013.

JAXA/NHK

First primate cells cloned

Researchers in the United States created embryonic stem-cells from cloned primate embryos, publishing their results in *Nature* in November. A similar success in humans could lead to therapeutic cloning. One researcher in the field said that the results were "like breaking the sound barrier". After the controversy surrounding Woo Suk Hwang's fraud, *Nature* took the unusual step of having the work independently verified.

Russia claims the North Pole

A previously obscure United Nations' rule that allows countries to claim parts of the sea floor beyond the continental shelf was thrust into the public arena after workers on a Russian submarine planted their national flag on the bottom underneath the North Pole in August. "The Arctic is Russian," said expedition leader Artur Chilingarov. Canada, Norway and Denmark disagreed.

NEWS

Climate deal agreed in Bali showdown

Insults, threats, tears and booing: the latest round of international climate talks made for an entertaining, if gruelling, two weeks in Nusa Dua, Indonesia. These talks may well be remembered for the bold stand that developing countries took against the United States in the push for consensus on how to move forward in negotiating a new international framework on climate change.

Some 10,000 delegates from nearly 190 nations finally agreed on Saturday to a 'Bali roadmap' that will guide negotiations up until the end of 2009, when they will have to decide on a regime to replace the Kyoto Protocol in 2012. But the path to agreement was rocky from the outset, as discussions over individual words in the draft document led to heated arguments, threats of trade sanctions or boycotts, and even tears.

By the middle of the second week, a deal had more or less been reached on some of the key issues that would enable developing nations — including those with budding economies — to reduce their emissions. The deal entailed providing compensation for reducing tropical deforestation, which accounts for some 20% of greenhouse gases.

That left two thorny issues on the table, both of which the United States objected to. As one of the world's largest greenhouse-gas emitters and the only rich nation not to have ratified Kyoto, the United States became noticeably isolated as developing and developed nations, led by the European Union (EU), stood strong on the need for richer countries to lead on climate change and to tackle it with emissions targets. In a year that has been punctuated with reports from the Intergovernmental Panel on Climate Change (IPCC) on the urgency of global warming, most delegations agreed that the roadmap should refer to the need for industrialized nations to slash emissions by 25–40% of 1990 levels by 2020.

The EU delegation, headed by commissioner Stavros Dimas, argued that a roadmap without a destination would be pointless. And Portuguese secretary of state for the environment, Humberto Rosa said: "It is crucial for us that we must have an idea where we are heading to — it's not only to science to show us the destination, but the destination must be consistent with the science."

But the United States countered that to include specific numbers would be to "prejudge the outcome" of the process. Following days of intensive negotiations, a compromise was reached by including a

J. SAMAD/AFP/GETTY

People from around the world campaigned outside the Bali conference centre.

Exchange rate hits US researchers

The weak dollar is affecting US researchers working abroad and threatens American involvement in flagship projects say physicists at the CERN laboratory. Grants paid in dollars to researchers in Europe are now worth substantially less than they were a year ago.

"It's hurting, and people are scrimping and making up for it in other ways," says Mike Tuts, a programme manager for the US collaborators on Atlas, a component of the Large Hadron Collider (LHC) at CERN, which will look for the Higgs boson, dubbed the God particle.

In January 2006, a US dollar bought about €0.83, in January 2007 that was €0.77. Today it is worth only €0.69. Although US researchers at CERN receive adjusted pay to reflect the cost of living, these corrections are not keeping up with the change in exchange rates. "It's already starting to hurt and it's making people nervous," says Joel Butler of the US section of the Compact Muon Solenoid experiment, another part of the LHC.

And it is not just the researchers who are suffering. Local costs incurred by US teams at CERN have risen significantly. Butler says that additional funding to deal with these expenses will be requested, but it is not guaranteed to be forthcoming. "In the end, if we don't get some relief, we will probably have to reduce the size of the community and do more work remotely," he says.

Similar problems may befall CERN itself, which straddles the border between France and Switzerland. Changes in exchange rates between the euro and the Swiss franc — between 2002 and 2006 — cost the LHC project 40 million Swiss francs (US$35 million), says Florian Sonnemann, head of resource planning and controlling at the facility. "This SFr40 million had to come from the contingency we have for this project," he says.

However, at the moment, it is the Americans who are really suffering. "We used a rather poor exchange rate in the budget planning," says Tuts. "Ultimately it will impact on the science." ■

Daniel Cressey

SNAPSHOT
Bumper cone crop

The tall conifers of America's Pacific Northwest are experiencing a peculiarly heavy cone-harvest. Ken Bible of the University of Washington in Seattle, who investigates the phenomenon, says: "In normal mast years [when most Douglas-fir cones are produced], we would see four or five cones per branch. Now we are seeing 30."

The cone-production cycle is thought to be driven in part by weather events during the development of tree buds and cones. Bible says that his team is looking closely at temperature, rain, humidity and wind data at several development stages to see if they can work out what caused the super mast.

It seems to be confined to that region. For example, white spruce in Canada are masting normally in Edmonton, Calgary and Banff, reports Jalene LaMontagne, at the University of Calgary in Alberta. ■

Emma Marris

CLIMATE CONFERENCE
Find all the news from Bali, plus archived material on the IPCC.
www.nature.com/news/specials/bali/index.html

footnote reference to the IPCC's fourth assessment report, without mention of the numerical range on reductions. Many feel that this concession amounts to sidelining the science, and risks narrowing the window of opportunity to avert dangerous climate change.

Amid all the drama, the real blows were reserved for a grand finale on Saturday, when worn-out delegates wrangled over one remaining issue — whether rich nations should provide "verifiable, measurable and reportable" technological aid to developing countries. Seemingly inspired by Al Gore's speech on Thursday urging delegates to sidestep the "obstructionist" United States, and following much booing and hissing at the US delegation, Kevin Conrad, Papua New Guinea's ambassador for climate change, stated: "If you cannot lead, leave it to the rest of us. Get out of the way." In an eleventh-hour turnaround, the United States conceded.

"Bali has delivered what it needed to do," says Yvo de Boer, the executive secretary of the UN convention on climate change, calling the agreement "ambitious, transparent, and flexible". Perhaps most importantly, it has succeeded in bringing what may still be the world's largest emitter back to the table. But whether the 'flexibility' that was required will ultimately provide a means of manoeuvring out of real emissions reductions remains to be seen. ■

Olive Heffernan

See Editorial, page 1127.

THE SEATTLE TIMES

Q&A: Siti Fadilah Supari

Indonesia has been hit by more human deaths from the H5N1 bird-flu virus than any other country, yet it refuses to share its virus samples with the World Health Organization (WHO). **Declan Butler** talks to Indonesia's health minister.

H. D. NAM/AFP/GETTY

Why is Indonesia withholding samples that could track the virus's evolution and help produce a vaccine?
Indonesia is open to international collaboration but this must be fair, transparent and equitable. The WHO's Global Influenza Surveillance Network system is obviously unfair and opaque. Samples shared become the property of the WHO collaborating centres in rich countries, where they are used to generate research papers, patents and to commercialize vaccines. But the developing countries that supply the samples do not share in these benefits. In the event of a pandemic, we also risk having no access to vaccines, or having to buy them at prices we cannot afford, despite the fact that the vaccines were developed using our samples.

The above mechanisms can lead to a vicious cycle, in which poor countries become poorer because they have diseases, and industrialized countries become richer at the expense of poor countries. As a consequence, there will be some people who will create diseases and take advantage of the situation. This is a form of neo-colonialism and neo-capitalism.

What would it take to end the deadlock?
Earlier this year, the WHO adopted a resolution accepting the need to ensure fairer sharing of benefits to help developed countries prepare for and respond to a pandemic. Last month in Geneva we agreed concrete steps towards this goal. The first is that the WHO will develop a tracking system for viruses we send abroad. The second recognizes that sharing must respect national laws, which means that Indonesia will be free to assert its rights over samples. We are willing to share viruses for research purposes but if a commercial company wants to create a vaccine from those samples, then it must negotiate with Indonesia, which has rights over the samples.

But the key outstanding concession Indonesia demands is that all shared virus samples be subject to a material transfer agreement clearly setting out such rights. We already share our virus-sequence data with the Global Initiative on Sharing Avian Influenza Data (GISAID), a body created last year. I consider GISAID has established a fair and transparent mechanism for regulating genetic-sequencing databases because it includes a material transfer agreement.

Why is Indonesia not as successful as its neighbours at controlling H5N1?
Vietnam, as a centralized socialist country, can get high compliance on national policies and so has succeeded, for example, in implementing rapid culling of birds. Thailand's monarchy is well respected, resulting again in good compliance. In contrast, Indonesia is in transition towards a decentralized democracy after three decades of authoritarian national rule. We are still on a learning curve, and compliance of the relatively independent regional authorities with national policies is often poor. Indonesia is made up of more than 17,000 islands, which again complicates compliance.

The cultures are also different. Birds play an important role in Indonesian culture — pet singing birds are considered signs of respectable households, for example. In rural communities, backyard farming is a major income source, and has been a key element in improving nutrition.

But bear in mind that Indonesia has a population of more than 200 million, so one needs to keep in perspective the total of 113 cases of H5N1 — I'd argue that we've been fairly successful in controlling the disease.

But scientists say that Indonesia is still not doing enough.
It is the international scientific community that delays the acquisition of critical knowledge. If they wish to have rapid results, why don't they come to Indonesia and work with our scientists here? We have the necessary facilities, such as biosafety level-3 secure labs and good Indonesian scientists. That would also put an end to disputes over virus sharing. ■

SIDELINES

SCORECARD

Newton day
Evolutionary biologist and staunch atheist Richard Dawkins suggests rebranding Christmas to celebrate Isaac Newton's birthday, which also falls on 25 December.

ON THE RECORD

"I told her to stop admiring young men in Ferraris."

The UK government's science adviser, David King, reiterates the advice he gave a young woman on how to combat global warming. The Ferrari Owner's Club has called the comment "unhelpful".

"Of course he's sexy ... [he's] like watching a beautiful racehorse run."

Actress Uma Thurman reveals her impression of environmental advocate and Nobel laureate Al Gore (both pictured, below).

OVERHYPED

Scientific dating
Dating is all in the genes according to a US company that began offering the world's first DNA-dating service last week. The system is based on a theory that people with differing immune systems are attractive to each other. For a mere US$1,995, hopeful mates can be genotyped for three different immune system molecules. The scheme is backed by a handful of peer-reviewed papers showing some evidence for attraction — but there is little evidence for the service's claims that matches will experience more orgasms and less infidelity.

Sources: The Washington Post, The Daily Telegraph, The New Statesman, Yahoo News, The Scientist

L. ASERUD/SCANPIX/REUTERS

STATE ART MUSEUM, NIZHNY NOVGOROD, RUSSIA/BRIDGEMAN ART LIBRARY

Physicists make ripples with their 'magic carpet'

Perfectly timed for pantomime season, a team of scientists has come up with instructions for how to make a flying carpet.

The magical device may owe more to Walt Disney than to *The Arabian Nights*, but it is not pure fantasy, according to Lakshminarayanan Mahadevan of Harvard University in Cambridge, Massachusetts, and his co-workers. The researchers have studied[1] the aerodynamics of a flexible, rippling sheet moving through a fluid, and find that it should be possible to make one that will stay aloft in air.

No such carpet is going to ferry people around, though. The researchers say that to stay afloat in air, a sheet measuring about 10 centimetres long and 0.1 millimetres thick would need to vibrate at about 10 hertz with an amplitude of about 0.25 millimetres. Making a heavier carpet 'fly' is not forbidden by the laws of physics. But the researchers say that their "computations and scaling laws suggest it will remain in the magical, mystical and virtual realm", as the engine driving the necessary vibrations would need to be so powerful.

The key to a magic carpet is to create uplift by making ripples that push against fluids such as air or water. If it is close to a horizontal surface, like a piece of foil settling down onto the floor, such rippling movements create a high pressure in the gap between the sheet and the floor. "As waves propagate along a flexible foil, they generate a fluid flow that leads to a pressure that lifts the foil, roughly balancing its weight," Mahadevan explains.

But as well as lifting it, the ripples can drive the foil forward — a trait required by any respectable magic carpet. "If the waves propagate from one edge," says Mahadevan, "this causes the foil to tilt ever so slightly and then move in one direction towards the edge that is slightly higher. Fluid is then squeezed from this end to the other, causing the sheet to progress like a submarine ray."

Speed bumps

To travel at speed, the carpet would have to undulate in big ripples, comparable to the size of the carpet. This would make the ride very bumpy. "If you want a smooth ride, you can generate a lot of small ripples," says Mahadevan. "But you'll be slower."

"It's cute, it's charming," says physicist Tom Witten at the University of Chicago in Illinois, who is intrigued that the researchers thought to study such an unusual engineering feat.

So could tiny flying carpets really be made? Spontaneous undulating motions have been demonstrated in 'smart' polymers suspended in fluids — the polymers can be made to swell or shrink in response to external signals. In September, another team at Harvard University described flexible polymer sheets coated with cultured rat muscle cells that flex in response to electrical signals and are able to swim[2].

"In air, it should be possible to make moving sheets — a kind of micro hovercraft — with very light materials, or with very powerful engines," Mahadevan says. ■

Philip Ball

1. Argentina, M., Skotheim, J. & Mahadevan, L. *Phys. Rev. Lett.* **99**, 224503 (2007).
2. Feinberg, A. W. *et al. Science* **317**, 1366–1370 (2007).

UK government claims half of MRC's patent profits

The UK Medical Research Council (MRC) is being forced to hand over half of the money it has earned from exploiting its intellectual property.

The MRC's £200-million (US$400-million) 'commercial fund' consists of the profits from discoveries such as monoclonal antibody technology. In the past, the MRC had been allowed to keep the money and use it to build medical research facilities.

But on 14 December MRC officials announced that £92 million would be given over to the government's consolidated fund, which is used by the Treasury to fund general programmes. Martin Rees, the president of the Royal Society, Britain's main independent scientific body, called the move a "breach of faith" with scientists.

"In fairness, a lot of the money is being retained for science," says Leszek Borysiewicz, the MRC's chief executive. The MRC will also be able to keep future intellectual-property earnings, as long as they are forecast correctly. But he adds that the decision by the Treasury is "worthy of debate".

Pasteur Institute fined over patent dispute

A French court has slammed the ethical behaviour of Paris's prestigious Pasteur Institute and ordered it to pay €150,000 (US$215,000) in damages to Rupert Mutzel, a biologist at the Free University of Berlin, for denigrating his scientific reputation.

In the 1990s, Mutzel and Philippe Marlière — then a researcher at the Pasteur Institute — patented a microbial technique for selecting accelerated proliferation of living cells in suspension. An exclusive licence was agreed for Marlière to utilize the patent. But in 2003, the institute revoked the licence without asking Mutzel, and negotiated a new one with another company, Eco-Solution.

The court ruling, which became public on 12 December, says that the institute tried to mislead the court by concocting "totally imaginary" evidence against Mutzel. The court also criticized the institute's ethics committee for throwing doubts on Mutzel's integrity and for saying that it "regretted" that he had "provoked a legal conflict". The court told the committee that it was normal for Mutzel to take legal action to defend his rights.

The Pasteur Institute says the current management does not dispute Mutzel's status but adds that it will appeal against the size of the fine. A separate case against the institute by Marlière is pending.

Congress tightens rules on fuel standards for cars

Democrats in the US Congress have struck a deal on legislation to increase the fuel efficiency standards for motor vehicles — the first such increase in more than three decades.

The legislation would increase the standards by roughly 40%, to 15 kilometres per litre, by 2020, while instituting a series of energy-efficiency requirements for appliances and federal buildings.

The bill would also mandate the use of 136 billion litres of biofuels annually by 2022, compared with roughly 25 billion litres this year. Of the total, 80 billion litres must come from advanced biofuels such as cellulosic ethanol and meet certain requirements for greenhouse-gas emissions.

Facing opposition from Republicans and the White House, Democrats had to jettison a federal mandate for the production of electricity from renewable sources and a tax package shifting energy subsidies from fossil fuels to renewable energies.

The Senate passed the bill on 13 December, and the House of Representatives was moving towards passage as *Nature* went to press this week. The president is expected to sign the bill.

CERN names its next head as collider is set to go online

CERN

Rolf-Dieter Heuer.

The next director-general of CERN, Europe's primary high-energy physics laboratory near Geneva, Switzerland, will be Rolf-Dieter Heuer, who will take the reins in January 2009.

Heuer, whose appointment was announced on 14 December by the CERN Council, is currently a research director at Germany's DESY laboratory in Hamburg. He will replace Robert Aymar, who has run the laboratory since 2004.

Heuer says his top priority will be to ensure a smooth transition from the construction of the Large Hadron Collider, a SFr10-billion (US$8.7-billion) particle accelerator to its operation after it is completed. Heuer also says that he will seek to improve cooperation with particle-physics laboratories in the United States and Japan.

He adds that he will "keep options open" for CERN's future, including its participation in the International Linear Collider, a proposed next generation of particle accelerator.

Science appoints Bruce Alberts as editor-in-chief

Biochemist Bruce Alberts, an American known best for his seminal textbook *Molecular Biology of the Cell* and an advocate of international scientific cooperation, will take over as editor-in-chief of the journal *Science* beginning in March.

Alberts served as president of the US National Academy of Sciences from 1993 to 2005, and is currently co-chair of the InterAcademy Council, an Amsterdam-based international collaboration of science academies from 15 countries. He is based at the University of California, San Francisco.

Alberts replaces Donald Kennedy, a biologist and former president of Stanford University in California, who has edited the journal since 2000.

Springtime outburst on Mars

NASA/JPL/UNIV. ARIZONA

Channels carved by vaporizing carbon dioxide splay across the south polar region of Mars in this image taken by the Mars Reconnaissance Orbiter.

Last week, at the American Geophysical Union meeting in San Francisco, project scientists reported that the channels probably form as the warmth of spring causes carbon dioxide in the soil to vaporize, then flow upwards and burst out like geysers. Some of the gas then rapidly freezes and falls back to the ground in a brilliant white frost pattern. The strange features, radiating out from the top of small hills, were snapped during the martian spring in March.

Small advances

Most people agree that the environmental and health effects of nanoparticles need a lot more study. **David Goldston** looks at why so little progress has been made.

For at least the past four years, experts have been calling, even pleading, for a tightly coordinated, well-funded US government research programme on the potential effects of nanotechnology on the environment and on human health and safety (see *Nature* **444,** 267–269; 2006). But despite those efforts, bolstered by support from Congress, the Bush administration has yet to produce even a strategic plan for how such a programme would work. Why?

The problem in this case is not ideological conflict, or indeed opposition of any kind. The nanotechnology issue may be unique in that the administration and both parties in Congress, as well as industry and environmentalists, all agree that a targeted research programme is essential if nanotechnology is to achieve its economic and social potential. They all acknowledge that materials behave differently at the nanoscale, and that science is not yet able to predict what kinds of harm nanoparticles might cause. They all fear that if research doesn't catch up with manufacturing soon — more than 300 products based on nanotechnology are already in the marketplace — sooner or later something bad is likely to happen, prompting a backlash that could stifle work in the field.

Ironically, the broad agreement over fundamentals may be part of the problem. With no grand debate over first principles and no accusations of acting in bad faith, nanotechnology has received only fitful attention from the media and has been the subject of only a few congressional hearings. And with no divisive fight over regulation (partly because agencies don't know enough yet to figure out how to regulate), nobody has been screaming for research either to settle a policy debate or to forestall action, as has been the case with climate change, for example.

As a result, the administration's efforts to organize the research have proceeded at a steady but workaday pace that has fallen behind even the White House's own deadlines. In 2003, the administration set up the interagency Nanotechnology Environmental and Health Implications Working Group, chaired by the Food and Drug Administration. So far, that group has produced two reports that list priority areas for research. The most recent list, released in August, enumerated 25 extraordinarily broad priorities — "evaluate abiotic and ecosystem-wide effects", for example — without ordering them further or assigning them to any agency. Further direction

PARTY OF ONE

is supposed to be provided in a strategic plan, now promised for early next year, but no one is sure how specific it will be.

A worthwhile strategic plan would clearly identify what questions regulatory agencies need answered — for example, what characteristics of particles would make them dangerous to humans if inhaled or ingested? At what concentrations could particles be released into the environment without endangering species or ecosystems? The plan would also need to have a fairly detailed description of the basic and applied research that should be done to answer the questions, and then assign priority levels, agencies and funding to that research.

In the meantime, the administration calculates that seven agencies spent about $48 million on nanotechnology environmental and health research during the fiscal year that ended on 30 September and has proposed increasing funding by about $11 million this year. But no one is sure how accurate that accounting is: each agency self-reports its own, largely uncoordinated efforts, using its own definitions. A study of fiscal year 2005 nanotechnology spending by the Woodrow Wilson International Center for Scholars concluded that only $11 million had been spent on work that was "highly relevant" to environmental and health issues, while the administration had reported spending almost $35 million in that area.

Moreover, non-government experts argue that even the official levels are woefully inadequate. At a 31 October hearing before the US House Committee on Science and Technology, witnesses from industry and the environmental movement argued that environmental and health research should amount to 10% of research spending on nanotechnology — admittedly, a somewhat arbitrary figure — which would have meant $135 million last year.

The state of play is unlikely to change unless Congress credibly threatens to allocate spending specifically for environmental and health research and to centralize authority for such research in a single agency or office. Under every administration, agencies object to such steps so, as Samuel Johnson said about the threat of a hanging, they may concentrate the mind.

The status quo will never result in an adequate programme because no agency views the research as a high enough priority, given pressures to address immediate problems and ongoing programmes with limited funds. In addition, agencies always resist coordination — not surprisingly, they fight for autonomy — so most interagency plans amount to little more than stapling together proposals that each agency has developed on its own. In this case there's yet another hurdle: basic-research agencies are suspicious of efforts that try to direct research to answer specific questions formulated by the government to address policy needs.

These agency inclinations to remain independent and to favour proposals that originate with researchers are generally healthy tendencies that can, for example, ward off undue political interference and foster diversity in research. But they are maladapted for situations that call for focused research to resolve societal issues that need to be faced with some urgency, such as those associated with nanotechnology.

In such cases, the White House or, failing that, Congress needs to cajole, prod and even mandate agencies to move forwards quickly, and in unison. No one has done that sufficiently in this case, which is why at the science committee hearing, Andrew Maynard of the Wilson centre described the "overall federal government response to identifying and managing nanotechnology risks" as "slow, badly conceptualized, poorly directed, uncoordinated and underfunded". Inertia, which is to say the standard interagency process, is not a strong enough force to overcome those failings, but that's what's been relied on so far. The plans for learning about the possible societal consequences of nanotechnology are simply too small. ■

David Goldston is a visiting lecturer at Princeton University's Woodrow Wilson School of Public and International Affairs.

BUSINESS

Regulator under siege

Science at the Food and Drug Administration is in need of a revamp, as **Meredith Wadman** reports.

The US Food and Drug Administration (FDA) has had a tough year. The Maryland-based agency, whose 10,000 staff are responsible for overseeing most aspects of food and drug safety, used to enjoy an enviable reputation as one of the world's most trusted regulators — especially in the field of drug safety, where many other nations simply follow its lead.

But in March, the agency's difficulties in ensuring food safety came under scrutiny after thousands of dogs and cats became ill and at least 16 died from eating pet food laced with melamine, an ingredient used in plastics. The chemical was traced to a Chinese-made ingredient in the pet food. And late last month, a high-powered panel of academics and industry managers delivered a stinging rebuke on the agency's scientific capability.

"The impact of the deficiency is profound precisely because science is at the heart of everything FDA does," says the report. "The agency will flounder and ultimately fail without a strong scientific foundation."

FDA: Science and Mission at Risk was written by a panel chaired by Gail Cassell, vice-president of research at drugmaker Eli Lilly, and delivered to the agency's scientific advisory board on 3 December. Agency commissioner Andrew von Eschenbach had asked for the assessment a year ago.

Fallback

The report claims that "American lives are at risk" because of what it terms "scientific deficiencies". It adds: "Not only can the agency not lead, it cannot even keep up with the advances in science," and calls for a doubling of congressional funding for the agency within two years.

The document is the latest in a long line of such diagnoses: last year, for example, the Institute of Medicine offered a stark assessment of the FDA's drug-safety capabilities (see *Nature* **443,** 372; 2006).

"If the agency had no new responsibilities, it would be inadequately funded," says William Hubbard, a former associate commissioner at the FDA. "When you add the fact that Congress is piling on new responsibilities and new workloads in things such as food imports, it makes the situation particularly dire."

Scott Gottlieb was the agency's deputy commissioner for policy until earlier this year and is now a fellow at the conservative American Enterprise Institute in Washington DC. He says that "the agency has probably reached a tipping point, partly because of the systemic lack of funding that has gone on for a very long time — but even more so because the complexity of its work and the risks inherent in the products it oversees have grown so much in the past few decades."

Funding for the agency has stagnated even though Congress has enacted 123 statutes over the past 20 years that add to its regulatory responsibilities. Almost all the statutes require scientific expertise, yet none has come with extra public money, and the number of FDA employees that congressional funds support is roughly the same as it was 15 years ago.

Drug industry 'user fees' have, however, substantially augmented the agency's income since they were introduced in 1993, taking pressure off Congress. In 2008, the fees are expected to exceed $460 million, but they are mostly targeted at drug and device review, and other functions have suffered since they were introduced. The operating budget of the FDA's Center for Food Safety and Applied Nutrition, for example, is a little over half of what it was in 2003.

Dire diagnosis

But the Pharmaceutical Research and Manufacturers of America (PhRMA), the drug industry's main lobby group, called the report "too grim". The report's "dire warnings about the scientific expertise of the FDA and its ability to do its demanding regulatory job would appear to be more than a little premature", says Alan Goldhammer, the group's deputy vice-president for regulatory affairs. Goldhammer notes that next year, user fees will augment the congressional budget of $1.6 billion by nearly 30%.

Some commentators who are wary of calls for government spending, such as Gottlieb, support the report's conclusions. Yet most observers fear that it is unlikely to have much immediate effect on a financially constrained Congress. And with the signing of a new, five-year user-fee law in September, that train for new FDA money has already left the station. "Congress has a long history of ignoring these kinds of reports," says Gottlieb, "especially when they assign some of the blame to Congress — which this report does."

On the ground at the FDA, the publication will probably do nothing to quell recruitment and retention problems, which continue to dog the agency. For some employees, it is already too late. "I felt that I could accomplish more with my skills and experience out in industry," says a former device reviewer, who added almost one-third to his salary when he left the FDA last year, "and that industry would appreciate me more." ■


DA/AP


L. V. BURKE/AP

Former FDA official William Hubbard (right) doesn't think the agency has the resources to adequately police food imports.

2007 GALLERY

IMAGES OF THE YEAR

Many formulations of the scientific method begin with observations. And the images here are indeed exciting observations — new pictures from Earth and space that will serve as the starting points for great science. But often the most arresting scientific images are captured at the end of the process, in the form of a solution. Such pictures represent the culmination of months of tireless work in the laboratory and have a still, completed quality. The structure of a protein or material, the high-resolution image revealing microscopic handiwork, the elegant visualization of data. Whether from the start or the end of an investigation — or from somewhere in between — these images are some of the most striking from 2007.

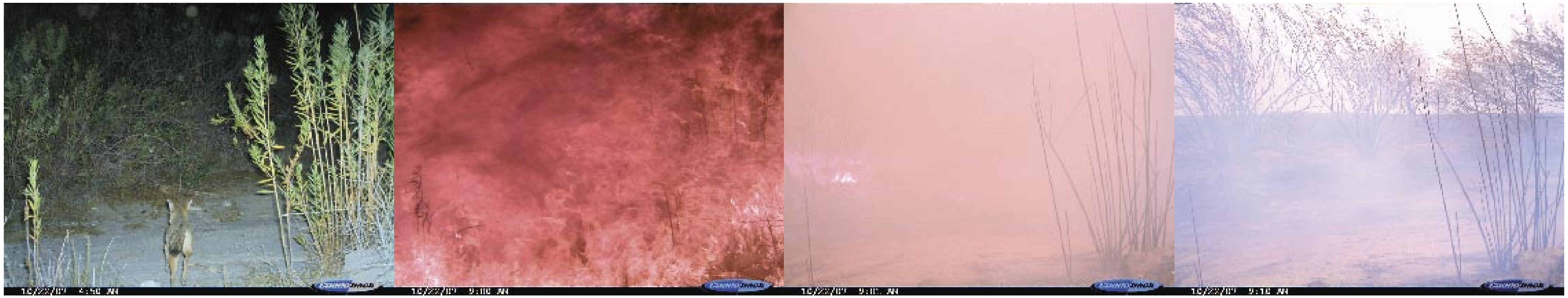
Researched and written by Emma Marris.

NASA/JPL/SPACE SCI. INST.

BACK OF IAPETUS

This false-colour mosaic is the first high-resolution image of the pale half of Iapetus, Saturn's two-faced moon. Iapetus is tidally locked, so it doesn't rotate, and it orbits Saturn with its dark face (just visible on the right) on its leading hemisphere. Cassini captured the picture on 10 September.

USGS



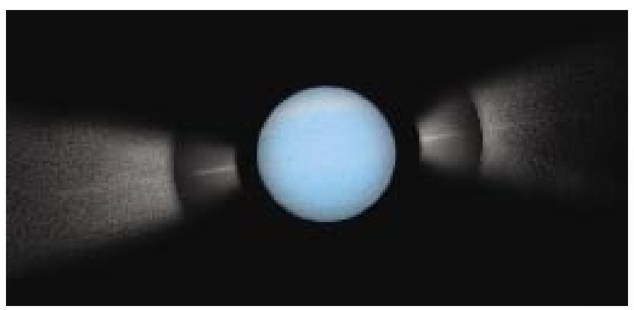
FIRE ON FILM

The Santiago fire in Orange County, one of the many wildfires in Southern California this year, was photographed by a camera set up to shoot wildlife. On 22 October, at 04:50, the camera snapped a coyote probably fleeing the fire. Then, at 09:00, the flames came. The final two pictures in the sequence were taken at 09:01 and 09:10.

NASA, ESA AND M. SHOWALTER/SETI INST.

RINGSIDE SEAT

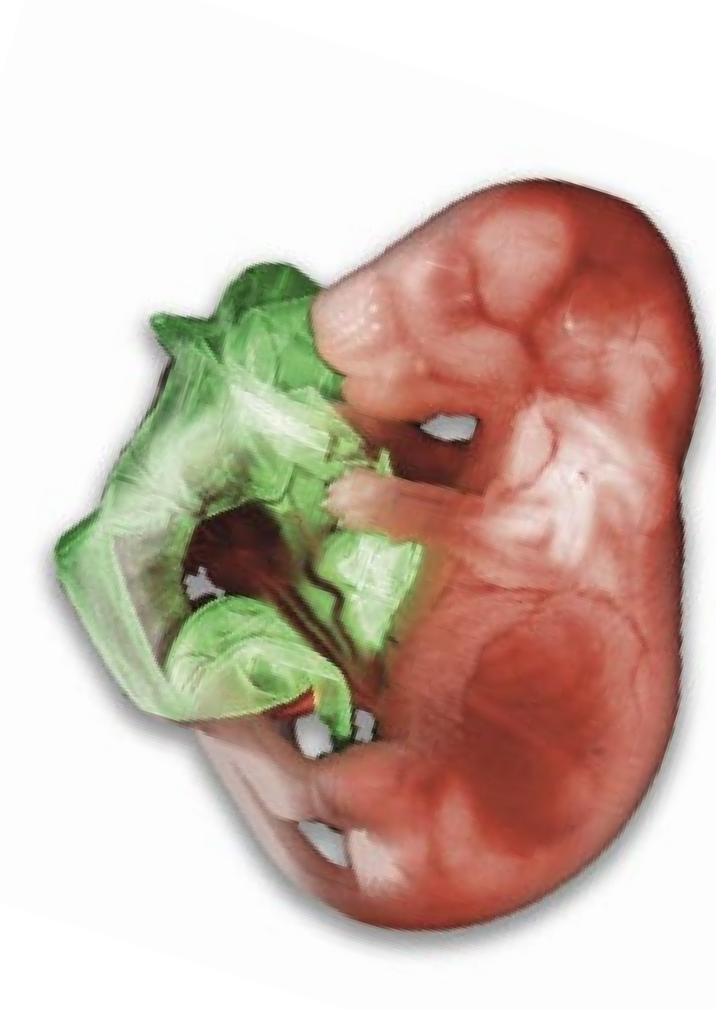
Uranus's rings are edge-on to Earth once every 42 years — and 42 years ago, no one knew the planet even had rings. So this year was the first chance to capture them from Earth, or at least its environs. The Hubble Space Telescope caught the moment on 14 August — in this enhanced image the rings show up as white spikes either side of the planet.



E. MASON

COMET LACE

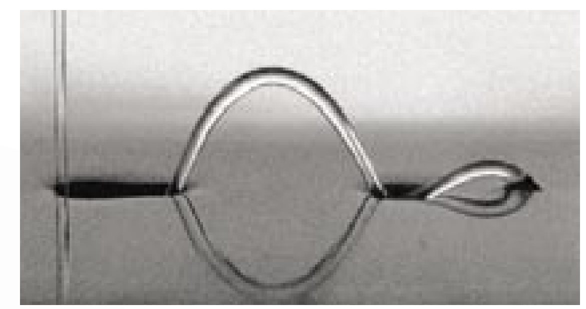
This is Comet McNaught as it steamed across the January sky, leaving a filigreed tail. It was the brightest comet to be seen from Earth for more than 30 years.



G. KWON, MEMORIAL SLOAN KETTERING INST./NIKON INSTRUMENTS

INSIDE THE EGG

This image of a transgenic mouse embryo and its yolk sac won first prize in Nikon's 2007 Small World competition. The yolk sac glows green and the mouse red because of different tagged proteins expressed in each.



SPLASHDOWN

This is a stream of silicone oil captured as it bounces twice before merging into an oil bath. The liquid jet, lubricated by a thin film of air, managed this leap because the oil bath was moving — in this image from left to right.

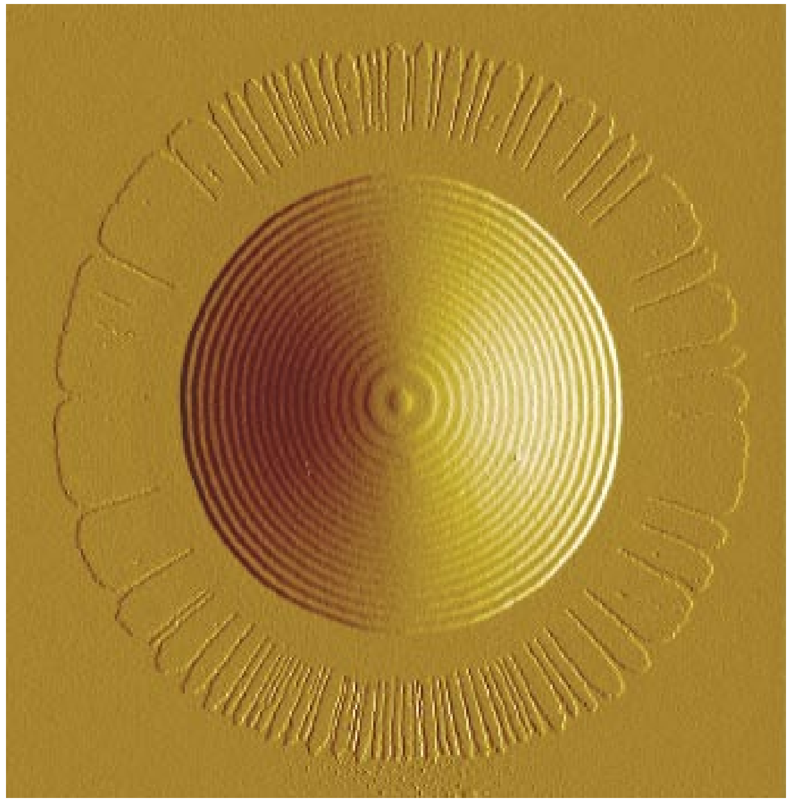
M. THRASHER, H. SWINNEY *ET AL.*

BEACH FOAMING

Beaches turned into bubble baths north of Sydney, Australia, this summer, when a perfect confluence of strong currents and myriad marine minutiae created mountains of foam.

B. COUNSELL/ICON IMAGES

AM. PHYS. SOC.

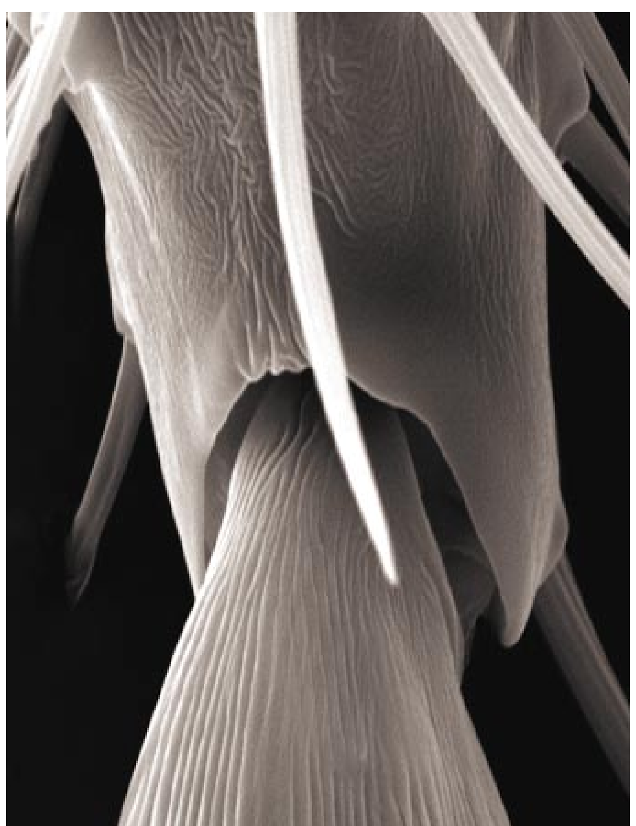


GOLD BLOSSOM

A nanoscale particle made up of 102 gold atoms is reconstructed here from crystal data. Many of the gold atoms collect in a central dodecahedron, but those arranged in the outer layers of the particle show some 'unanticipated geometries'.

DAZZLING DROPLET

Not all liquids form a dull dome when they sit as a droplet on an impermeable surface. This complex liquid of diblock co-polymers rests on silicon in architectural splendour.



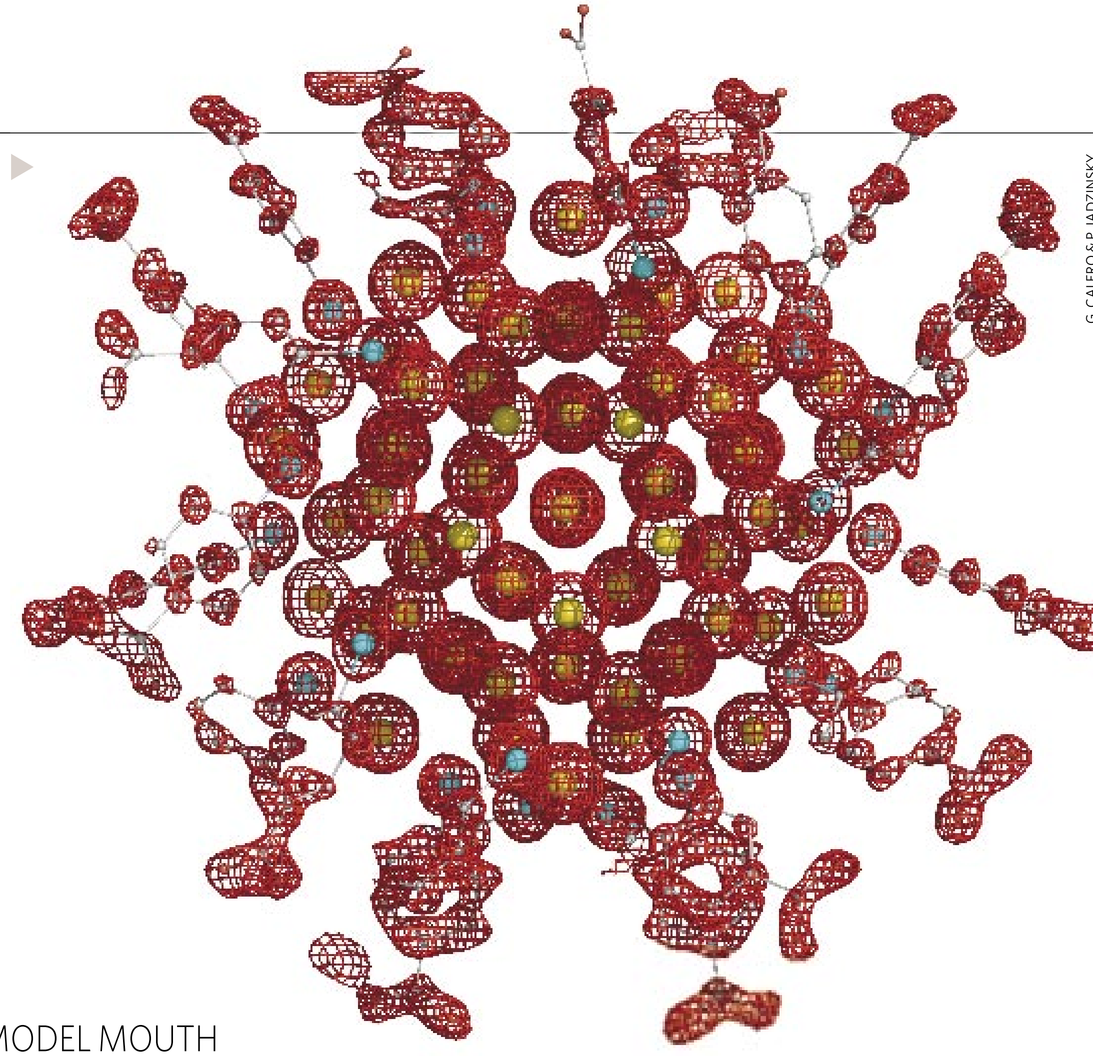
CARVED LEG

This picture of a fruitfly's leg joint was taken as part of a project to determine how the fly sculpts its limbs. The chasm inside the joint is carved out by apoptosis associated with a chemical gradient of a molecule called decapentaplegic.

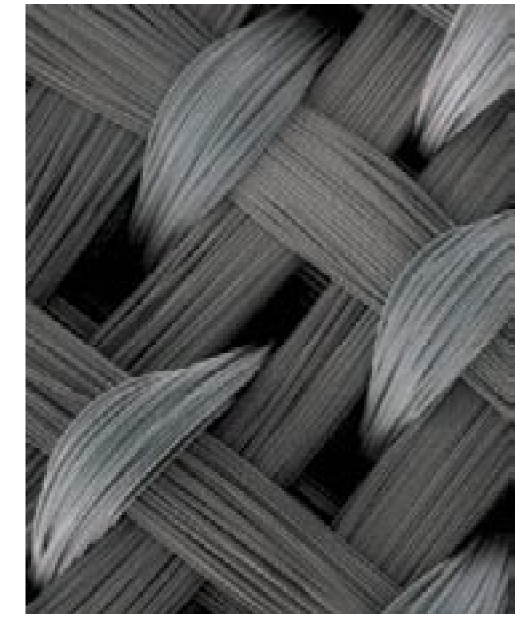
MODEL MOUTH

The stickleback fish, often used as a model in evolutionary work, has never looked so fierce. This image was taken as part of research into patterns of wear on teeth, which can reveal both what this fish has eaten and the dining habits of its fossilized ancestors.

M. PURNELL, UNIV. LEICESTER



G. CALERO & P. JADZINSKY



TINY LOOM

This small piece of three-dimensional weaving demonstrates a technique for making scaffolds that move and act like cartilage to help tissues to recover after trauma. The resulting fabric is about 1 millimetre thick and is woven from a yarn of polyglycolic acid.

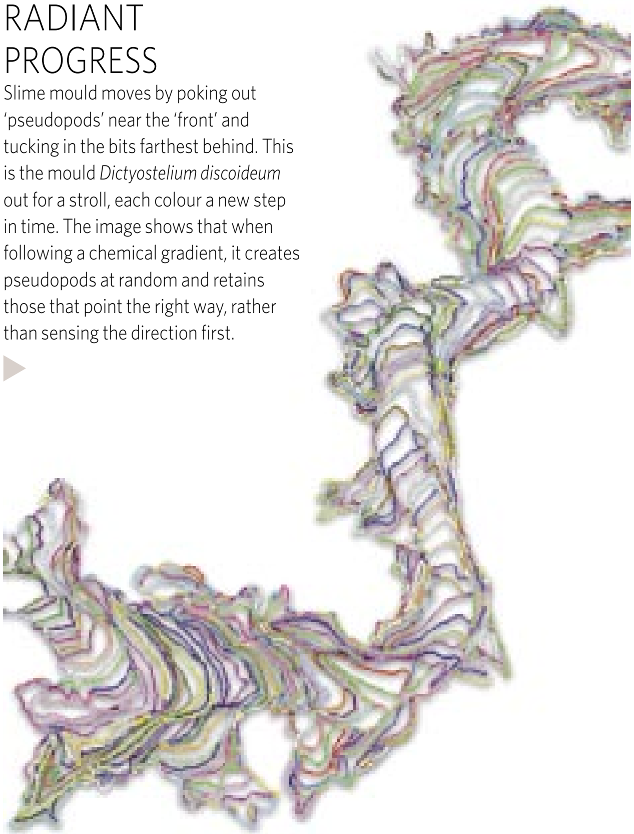
A. TKACHEV/ITAR-TASS

ANCIENT OFFSPRING

This 4-month-old woolly mammoth dates back some 10,000 years and was found in the melting Siberian permafrost. It was named Lyuba after the wife of the reindeer breeder who discovered her.

RADIANT PROGRESS

Slime mould moves by poking out 'pseudopods' near the 'front' and tucking in the bits farthest behind. This is the mould *Dictyostelium discoideum* out for a stroll, each colour a new step in time. The image shows that when following a chemical gradient, it creates pseudopods at random and retains those that point the right way, rather than sensing the direction first.



NEURONS IN GLAD RAGS

These motor neurons are sporting this season's colours, thanks to a labelling technique that lets chance mix up each cell's expression of four different fluorescent proteins (although only three proteins were used for this image).

NEOLITHIC EMBRACE

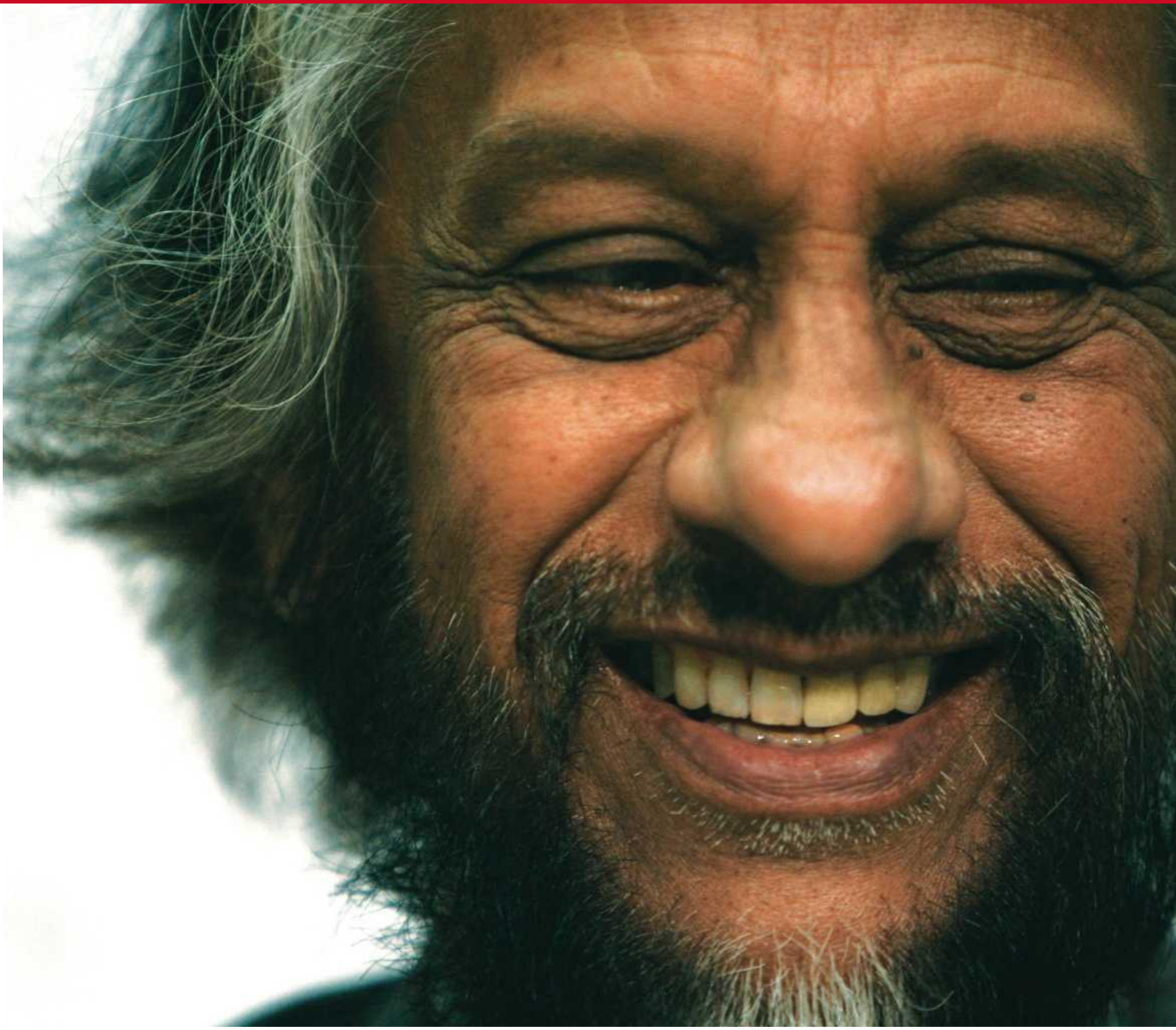
These roughly 5,000-year-old remains of a young man and woman were exposed in Valdaro, Italy, in early February, during an archaeological dig to pave way for the construction of a factory.

AP

M. BIUW

SCIENTIFIC CURIOSITY KNOWS NO BOUNDS

A southern elephant seal (*Mirounga leonina*) retains its dignity, despite having the latest in multi-parameter recording sensors stuck to its head.



The team captain

In 2007, the Intergovernmental Panel on Climate Change delivered its most thorough and authoritative assessment yet — and shared a Nobel prize for its efforts. **Gabrielle Walker** profiles its indefatigable leader.

R. SPRICH/REUTERS

RAJENDRA PACHAURI is standing quietly at the front of the small, plain room, his hands clasped behind his back. His expression is thoughtful as he weighs the words of his colleague, M. K. Halpeth. But look at his shoes — unexpectedly large, like a policeman's — and the sense of calm is undercut. Again and again he shifts his weight from foot to foot, or rocks backwards and forwards, heel to toe, toe to heel.

Several dozen of his employees are sitting in front of him in the meeting room. In his role as director-general of TERI, the Indian energy-research institute he has run for 26 years, Pachauri had come to the organization's Southern Regional Centre in Bangalore to present some senior members of staff with long-service awards. Three handsome gold trophies are lined up beside him on a wooden table. But before he could hand them over, his employees sprang a surprise on him, forcing him to wait and watch. Heel and toe.

When the Intergovernmental Panel on Climate Change (IPCC), which Pachauri chairs, was awarded a half share in the Nobel Peace Prize just a few weeks earlier, TERI's headquarters back in Delhi exploded into celebration with bottles of champagne — a waste to spray it, Pachauri admonished — sticky Indian sweets and the beat of traditional drummers. Now the staff at TERI's southern outpost wants its chance to fuss over him. Halpeth, the centre's director, is singing Pachauri's praises and presenting him with an enormous bouquet resplendent with irises, carnations and bright orange lilies. "You have made all Indians proud," he says.

Pachauri accepts the bouquet with grace, but immediately deflects the compliment. "I just happened to be captain of the team when this thing happened," he says. "And, as in cricket, the captain often gets credit for what the players have actually done." Then he launches into a stern call to arms. He has been looking at the figures, and the centre will struggle to meet its targets in the next quarter. "In the next couple of months you will need to go flat out," he tells them. And he urges them to think big, to think out of the box, to read all the time, to sharpen their intellectual skills. "The solutions that worked yesterday won't be adequate tomorrow."

On 10 December, a few weeks after his brief visit to Bangalore, Pachauri was on a far bigger stage in Oslo. Rather than giving the prizes he was receiving one — an award for the thousands of scientists whose team he has captained over its most recent five-year innings. And he presented a similarly urgent call to arms. Dealing with climate change means finding a path to development that is truly sustainable, he said. "We need to commit ourselves to this path before it is too late." Former US vice-president Al Gore, winner of the other half of the prize, echoed Pachauri's message. "We are what is wrong," he said, "and we must make it right."

On his way to collect his own medal, Gore stopped to shake Pachauri warmly by the hand. Patchy and Al, as they call each other, get along famously. It is all a far cry from the situation in 2002, when Pachauri beat Gore's favoured candidate to run the IPCC in a bitterly fought contest. Immediately afterwards, Gore lambasted Pachauri in the pages of *The New York Times* as the "let's drag our feet candidate", a patsy put in place to weaken the IPCC as one of various "acts of sabotage" by the new Bush administration. Pachauri had fought back with a letter of his own to the *Times*. "In a 1991 speech, Mr Gore [referred] to my 'commitment', 'vision' and 'dedication' … Will the real Al Gore please stand up?"

"He thought I was part of some kind of plot," Pachauri says. "Maybe he believed I had some sort of deal with the US administration, that I'd be soft in pushing the truth on climate change." If so, he knows better now.

****

It's not hard to see why an administration hostile to the science of climate change might have got the wrong impression of Pachauri. In background, international standing, scientific accomplishment and pugnacity he is a long way from Robert Watson, the IPCC chair whose bid for re-election Pachauri defeated. Pachauri is far from confrontational. He speaks softly and with old-fashioned courtesy. His dark beard, with its characteristic white stripe down the centre, gives him an ascetic air. He seems mild, self-effacing, even shy. But there is also that underlying sense of urgency that bears no resemblance to Gore's erstwhile accusations. At 67, Pachauri's feet do everything but drag.

At TERI — the initials stand for The Energy and Resources Institute — Pachauri's work habits are legendary. He can get off an international flight at two in the morning and go straight to the office. His driver says that on the 20-hour drive to TERI's Himalayan Centre in Mukteshwar, Pachauri never sleeps. He sits to attention in the back of the car, fielding phone calls, reading, thinking, sometimes chatting about the cricket.

He doesn't keep his work ethic to himself. Everyone at TERI is expected to show just as much dedication. Arrive even a few minutes late and you are likely to be greeted with a dry "good afternoon", whether you clocked off at six or at midnight. Do it several days in a row and you will receive an e-mail from the director-general reminding you of the values of hard work and discipline. And his staff love him for it. He was a hero to his employees long before the rest of the world took note. Although he has begged them to call him by his name, he is always "Sir", even when he's not in the room.

Rajendra Kumar Pachauri was born in Nainital, a beautiful hill station in the northern lake district of Uttarakhand. His father was an educational psychologist who earned his PhD in London before returning to work in his beloved India. His mother, born and educated in Burma (now Myanmar), never, ever, left a task unfinished; she insisted on the same from her children.

As a boy, Pachauri would sit for hours on the window ledge, gazing dreamily at the exquisite lake surrounded by mountains. His first poem, written at the age of eight, was published by *The Illustrated Weekly of India*. His mother expected him, the most sensitive of her three sons, to find a respectable position in the civil service. But when Pachauri attended the élite La Martinière College in Lucknow his career took an unexpected turn.

It was there that he met Arthur Flynn, a stickler for attention whom the students called "Mamu", which means 'mother's brother'. If you were distracted from one of his maths lessons, a well-aimed piece of chalk would bring you back to Earth. Like Pachauri's mother, Mamu demanded the highest possible standards. Pachauri loved his insistence on intellectual rigour, and discovered that he loved maths too. In the winter holidays, Mamu sent him home with a book of extra problems to work through. Pachauri found each solution incredibly satisfying; the work set him on a course that would lead to engineering.

Arthur Flynn is now in his mid-80s and suffering from Alzheimer's. He lives with his wife and daughter in Bangalore. Encouraged by an article Pachauri wrote in 2006, titled "My best teacher", for *The Times of India*, the family got in

touch after the Nobel announcement. Would Dr Pachauri like to visit?

Pachauri immediately set about figuring out how to fit a visit into his perpetually crowded schedule. On his hectic six-hour trip to Bangalore in November he reckoned that he could cram Arthur Flynn in between his lunch meeting with multibillionaire software magnate Azim Premji and the prizegiving and other duties at TERI's Southern Regional Centre. This was optimistic. Flynn lives on the other side of town, and Bangalore's notorious traffic was even worse than usual. Pachauri sat out the traffic jams with gift-wrapped parcels (a tie for Flynn, chocolates for his wife) beside him on the back seat of the car. He was both eager and hesitant.

When he arrived at Flynn's apartment the old man was sitting on the sofa, unable to rise. His family hovered excitedly. Flynn gathered that this visitor was important and whispered: "I want to ask his blessing."

"Papa, you're his teacher," said his daughter. "He is here to ask your blessing."

Pachauri knelt down and touched Flynn's feet. The old man rested his hands briefly on Pachauri's bowed head.

"I owe it all to you," Pachauri said.

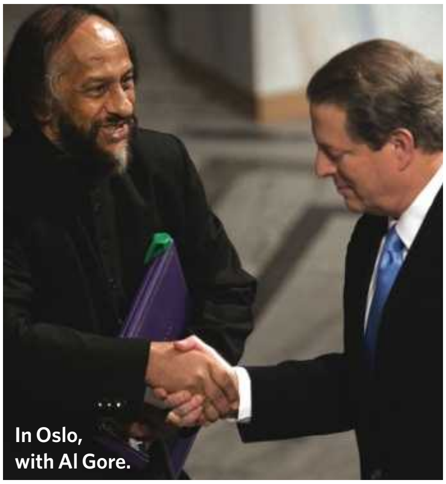
A few minutes later, just long enough for a sip of tea, a samosa and some sweets, and promising to visit again, Pachauri is hurrying back to business. But he stops just beyond the door and gives a deep, satisfied sigh. "It enriches the soul to make these connections," he says. Why? Did he feel that he had repaid some sort of debt? Pachauri hesitates. "Well yes," he says, "I owe a great deal to him." But it's obviously more than that. "Just to talk to him gives me a high," he goes on. "Human relationships are all that life is about. And there's something wonderful about the relationship between a teacher and a student. This was an opportunity to delve inside and get that richness of feeling that otherwise stays submerged. It's an experience that goes deep."

The spiritual sustenance remains with him; the serenity soon departs. "Look at the state of our traffic!" he complains as the car struggles back. "All of this should have been anticipated years ago. But our system is so hopeless at looking beyond the immediate interests of the individual." Does he feel hopeless? "No. Not at all. I just feel that we need to work harder and harder." Then he laughs. "That's why I'm a slave driver."

As it happens, that's the same phrase that one of TERI's executive directors, Leena Srivastava, uses when describing Pachauri. She should know — she has worked with him there almost since he took over as director in 1981. When he left school, Pachauri worked for India's railways, but in the 1970s he went to the United States, where he took two doctorates, one in engineering and another in economics, at North Carolina State University in Raleigh. His wife wanted to stay there but, like his father before him, Pachauri felt the tug of home.

A few years after returning, he accepted the directorship at TERI. Back then, the 'T' stood for Tata — the business group that had set up the institute mainly to provide funding to external researchers. The board wanted TERI to become a serious research institute in its own right, and persuaded Pachauri to oversee the transition. The plan was that the research would take place in Mumbai, near the business's headquarters. Pachauri, though, wanted it to be in Delhi, where it might attract intellectuals and influence the government. While Tata officials tempted him with talk

"It's one thing to make sure that our reports are sanitized. It's another for me as an individual to talk about policies that might work."
— Rajendra Pachauri


In Oslo, with Al Gore.

O. ANDERSEN/AP

of a beautiful Mumbai flat and spacious offices, Pachauri set up shop in a Delhi guest house. The kitchen and dining area became his office. He engaged a part-time secretary.

Today, TERI's staff runs to more than 700 and there are TERI offices in London, Washington DC, Tokyo, Dubai and Petaling Jaya in Malaysia as well as the four regional centres in India at Bangalore, Guwahati, Goa and Mukteshwar. The heart of the operation is the Delhi headquarters, now housed in a seven-storey building in the magnificent India Habitat Centre. No longer part of Tata, TERI is an independent non-profit organization, part think-tank, part research institute. Employees seek out their own projects or put in bids for international tenders, often in collaboration with institutes from other countries. They study the effects of climate change, possible adaptations, new energy technologies and their implementation in rural areas. They also investigate ventures with the potential to be part of the Kyoto Protocol's Clean Development Mechanism, through which developed nations can meet part of their commitments to reduce greenhouse-gas emissions by investing in energy projects in the developing world.

TERI's clients include Britain's Department for International Development, the World Bank, the Canadian International Development Agency and the Indian government. They also include major industrial companies; TERI sees its links with industry as a major strength. A current project with BP is investigating the potential to generate biodiesel from jatropha, a traditional oil crop — and by so doing to help farmers and to reclaim marginal lands (see *Nature* **449,** 652–655; 2007). "We are grounded in the real world," Srivastava says. "You can't work in abstracts. You have to take climate change to the people."

This same deeply practical streak extends to the complex that Pachauri insisted on building in the outskirts of Delhi. The RETREAT — Resource-Efficient TERI Retreat for Environmental Awareness and Training — sits among the farmhouses owned by Delhi's élite, just a few miles short of the burgeoning energy-hungry high rises that house the workers for many of India's call centres. It's designed to embody sustainable ideals. Delicate trellises called jalis, copied from ancient forts, allow breezes to pass through the buildings' walls. Solar-panel mosaics sit on the roof. Solar chimneys draw air up from tunnels below, cooling the rooms in summer and warming them in winter. It makes a great conference centre. Indeed his friend Al's people have recently checked it out as a potential venue for a meeting.

Pachauri loves going to the retreat. He says he'd like to live there. He especially relishes its two cricket grounds — the "Patchy greens", as the staff calls them. Hundreds of cricketing trophies take pride of place in his office in Delhi, and the corporate team tries to schedule TERI's matches for whenever Pachauri is in town. He is not just a spectator — he is a handy swing bowler. He has taken 348 wickets for TERI's team and is anticipating the 350th with boyish enthusiasm.

Cricket got into his blood when he was seven, and it has been there ever since. "It wouldn't be enough for him to go into coaching," says a team mate. "He wants to play and he plays every match to win." He never spares himself when it comes to hurling himself at the crease or stopping a fast ball; indeed, in 2005 a cricketing injury was to make one of his hardest tasks as IPCC chairman harder still.

In the run-up to the IPCC plenary at Addis Ababa, Ethiopia,

in April 2005, Pachauri was desperately concerned that some countries were mounting a campaign to remove the fourth assessment's 'synthesis report' — the summary of summaries that, unlike the massive working-group reports it draws on, stands a chance of being read by non-experts. Then, just before the meeting, when he threw himself with typical vigour at the crease to avoid being run out, he tore off his cricket pads and ripped the muscles of his leg. Dry ice that was supposed to help instead burned the skin; his thigh swelled up alarmingly. His family begged him not to fly, but he insisted. In Addis he limped painfully through the sessions, persuading, arguing and building a consensus with his own brand of polite implacability. Finally, the dissenting countries said they would agree to a synthesis report if it were short; Pachauri took what he could get. And it wasn't long before he was back at the wicket.

Pachauri is true to his boyhood in another way, too: writing poetry. In 1993 he published a slim volume called *Moods and Musings*, which was co-written with his daughter Rashmi Pachauri-Rajan, and dedicated to his mother. The poems are conventionally styled. Love and its regrets loom large — but in some a more angry passion wells up, as when he channels the feelings of a television viewer watching the suffering of a boy and his mother in Bangladesh:

'Ma, Ma' he sobs
To the tinkle of his melancholy bell. That meagre
Handful of uncooked rice, she gives him
Till weary of chewing, he slips into slumber's security.
As I sip Chivas and soda, I pause to laugh at Tony Randall —
A commercial they show to extol the purity
Of dog's food with 'beef chunks, vitamins and minerals
For a complete and balanced diet.' The man says
'Doesn't your dog deserve Alpo?'

The poems are another way for Pachauri to delve into himself; many, probably most, are not destined for publication. "You just feel an urge to write it and get it out of your system," he says.

****

Pachauri joined the IPCC as a lead author for its second assessment report, published in 1995. By the time of the third assessment, he was serving as one of the five vice-chairs. After that, in 2002, he decided to stand for the top job. He's an ambitious man, as the growth of TERI shows. "One is always looking for a bigger challenge in life," he says, still his mother's son.

It was in some ways an unlikely candidacy. The organization's two previous chairs had both been physical scientists: Bert Bolin was a meteorologist, Robert Watson an atmospheric chemist. They had both been highly regarded academically, and had worked for international organizations, Bolin as scientific director of the European Space Research Organization (a precursor to today's European Space Agency), Watson as chief scientist at the World Bank. Pachauri, on the other hand, was an engineer and economist by background, an administrator by experience. He was not very well known outside India. More damningly, in the eyes of some, he and TERI had close relationships with energy companies. He was even on the board of the Indian Oil Corporation. Worse still, Pachauri came to be heavily supported by a new US administration, led by President George W. Bush, that was sceptical of climate change.

The IPCC, which was set up as part of the United Nations Framework Convention on Climate Change, works by assembling teams of scientists willing to synthesize the current state of knowledge about climate change, its causes and its effects into periodic assessment reports. Watson, who had presided over the third assessment report, published in 2001, was eager to reprise his role. But the forthrightness that made him popular among climate scientists did not serve him well with the industrial carbon lobby, or with a US administration that was close to that lobby and wary of Watson's close links to the previous Clinton administration. On 6 February 2001, a memo from ExxonMobil to the White House (which was later leaked) asked "Can Watson be replaced now at the request of the US?"

"If anything, Patchy has spoken out more forcefully than I did." — Robert Watson

The United States threw its weight behind Pachauri, who had been nominated for the chair by India. Many developing countries lent their support. Watson was distraught. He proposed that he and Pachauri should be co-chairs. "I still think we would have made an excellent team," he says now. But Pachauri would have none of it. "Two co-chairs is an unworkable concept except for someone who is desperate to keep the title of chairman in any form," he told *Science* at the time, although he now says he regrets saying something so intemperate. In the end, the vote went in Pachauri's favour, 76 to 49. Al Gore, still smarting from a previous electoral loss, wrote his angry letter to *The New York Times*, and the world settled down to see how Pachauri would fare.


In Delhi, at the office.

R. RAI/MAGNUM PHOTOS

"It wasn't a pleasant experience at all," says Pachauri. "But having got into it I didn't see why I should back off." Besides, he felt and still feels very passionately about what he refers to constantly as "the cause".

"We have been so drunk with this desire to produce and consume more and more whatever the cost to the environment that we're on a totally unsustainable path," he says. "I am not going to rest easy until I have articulated in every possible forum the need to bring about major structural changes in economic growth and development. That's the real issue. Climate change is just a part of it.

"If you want to reduce the emissions of greenhouse gases you have to tackle the structure of every sector of the economy. I think there's no point in concealing this fact. In fact the sooner we get it out in the open the better. This doesn't mean that we're advocating sackcloth and ashes or that we go back to a pastoral form of existence or live in caves. But we have the technological and economic muscle to transform the system that we have today into something that is less burdensome to the environment, less threatening for the conservation of natural resources and that lessens the global inequities in income and wealth."

These are not the words of an oil-industry stooge or a Bush placeman. "If anything, Patchy has spoken out more

forcefully than I did," says Watson. And this has caused some trouble. The IPCC's mandate is to be 'neutral with respect to policy' — to set out the options and let policy-makers decide how to act. The reports themselves reflect this. Every word is checked and double-checked by scientists, reviewers and then government representatives — "sanitized", as Pachauri puts it. But Pachauri is the face of the IPCC, and he often can't resist speaking out, despite a few "raps on the knuckles" for his comments. He insists that he always makes it clear he is speaking on his own behalf and not for the IPCC. "It's one thing to make sure that our reports are sanitized. It's another for me as an individual to talk about policies that might work. I feel I have responsibility far beyond being a spokesman for the IPCC. If I feel there are certain actions that can help us meet this challenge, I feel I should articulate them."

"I think Patchy needs to be careful," says Bert Metz, a senior researcher at the Netherlands Environmental Assessment Agency in Bilthoven, who is one of the co-chairs of the IPCC's working group on greenhouse-gas mitigation. "One of the things about the IPCC is that it lays down the facts. If you start mixing [that] with your own views that's not very wise. But he gets away with it because of his charm." Steve Rayner, director of the James Martin Institute at the University of Oxford, UK, and a senior author with the same working group, feels that Pachauri's personal statements place too much stress on lifestyles and not enough on technologies. But he also concedes that a certain amount of outspokenness is an essential part of the job. "I don't think you can provide inspirational leadership in an enterprise like this unless you are passionate. That's something Bob [Watson] and Patchy have in common. They are both very passionate about the issue and I think that's appropriate."

In other ways, Pachauri turned out to be Watson's opposite. Rather than confrontational, Pachauri's leadership style is heavily consensual. He gives the protagonists time and space to come to their own agreements. Sometimes a lot of time. Pachauri will patiently go through the night if he feels that's what is needed. Take last month's meeting in Valencia, Spain, to agree the text of the synthesis report that Pachauri had battled for in Addis Ababa. Pachauri had been alarmed, as the week of meetings got under way, to see unusually high levels of representation from the oil-exporting nations. "I thought, 'they have lined up their warriors.'"

Pachauri and the synthesis report's core writing team had spent the previous weekend going through all the comments they had received on a draft report and anticipating the problems to come. During the plenary itself, Pachauri set up contact groups on the issues that he thought would cause the biggest trouble: the report's "five reasons for concern", which set out the biggest dangers to come; the section on key vulnerabilities; and the explicit statements of how impacts would become more serious as temperatures rise. The meetings went on each evening until 10 or 11 at night.

By Thursday, with a host of objections still to work through, Pachauri took his session right through the night. When interpretation facilities stopped at 2 o'clock on Friday morning Pachauri persuaded the delegates to continue without them, using English as a working language.

Friday came and went; objections were still coming in after sunset. A press conference was scheduled for the next morning at which UN secretary-general Ban Ki-moon was due to speak. Pachauri had made the invitation personally; he was determined that the report would be ready. Around 7:30 that evening he finally showed his annoyance. "I haven't slept for 40 hours," he told the delegates before calling a brief

"Human relationships are all that life is about."
— Rajendra Pachauri


In Valencia, with Spanish deputy prime minister María Teresa Fernández de la Vega.

break as a cooling off period. "And I can go another 40 hours if necessary. But what message are we sending to the world?" The intervention worked — after the break the remaining objections evaporated. The next morning, Ban Ki-moon's emotional comments about his personal experience of climate change were beamed around the world, and the last words of the fourth assessment report were set before the planet whose future they described.

****

For all his diplomacy and determination, Pachauri did not put his own intellectual stamp on the fourth report as Watson had on the third. Although Pachauri will speak out at public events, his consensual style does not lend itself to forceful leadership in intellectual discussion or the shaping of ideas. "He tends to let people do the talking until they find their own consensus," says Metz. "You can end up waiting quite some time, which can be frustrating." When the lead authors of the different working groups disagreed — as they did on various things, most acrimoniously on the question of whether the natural-science assessment by Working Group I should look at scenarios in which there was deliberate action to mitigate climate change — Pachauri was unwilling to take one side or the other. Instead he stayed cool while temperatures rose around him. At least one of the senior scientists who unsuccessfully campaigned for mitigation scenarios to be assessed still feels let down.

And then there's the synthesis report as agreed in Valencia. The IPCC's three working-group reports are crafted in large part by the relevant group's two co-chairs. But the synthesis report is the responsibility of the overall chair. Watson describes writing the synthesis report for the third assessment as "the best fun I ever had in my life". The fourth assessment's synthesis report lacked that added value that Watson had relished supplying with a handpicked team; it was more a simple retelling of the salient points.

D. TUSON/AFP/GETTY IMAGES

"The new synthesis report may have been a bit more of a cut-and-paste job," Watson says, "but that wasn't Patchy's fault." Instead, Watson lays the blame with the format into which the synthesis report was forced at the Addis plenary and the sheer difficulty of getting the countries to agree even on that.

Pachauri's procedural victories and persistencies are his great contribution; the report's reception as the IPCC's most impressive yet bears witness to them. Still, the report belongs to the group as a whole. "The IPCC is far more than any individual," says Watson. "When Burt [Bolin] chaired, it was teamwork. When I chaired it was teamwork. When Patchy chaired, it was teamwork. It takes all the chairs, all the authors, all the reviewers, everyone. That's why I was so pleased by what the Nobel committee said. They didn't give it to Burt or me or Patchy. They gave it to the IPCC over the 20 years of its existence, and that's the strength of the thing." Pachauri invited Bolin, as founding chairman, to accept the Nobel medal on the IPCC's behalf, but Bolin's poor health prevented him from doing so.

Pachauri's concerns now are to oversee whatever reshaping of the IPCC its members see as necessary — there is much discussion of replacing or augmenting the mammoth assessments with more focused projects that are turned around much more quickly. Pachauri has not yet said whether he will be seeking another term as chair when the post comes open for re-election — "You know, it takes a lot out of you" — but he does say that he wants to focus more energy on his work at TERI and on serving 'the cause' by spreading the message. Speaking to a crowd of Bangalore intellectuals during his lightning visit to TERI's Southern Regional Centre, Pachauri warns them that the state of the world is even worse than he had feared, then fields a storm of questions.


In Delhi, on the pitch.

TERI

How can we deal with the population problem? Educate girls. What does he think of the Kyoto Protocol? "It's a beginning, but it's certainly inadequate. Let's hope that beyond 2012 we see much more effective action." Do the developed countries feel ashamed of what they have done? "Some do. But I have also been told that the American standard of living is 'non-negotiable'." Why shouldn't Indians aspire to the same standard of living as the Western world? "Gandhi was asked if he wanted India to reach the same level of prosperity as the United Kingdom. He replied: 'It took Britain half the resources of the planet to reach its level of prosperity. How many planets would India require?'"

Elsewhere in the world, Pachauri has been adamant that India has a long way to develop before it can consider taking on an emissions-reduction target. But here at home he warns constantly that India and the other developing nations can't afford to follow the path that the industrialized nations took.

"It's all very well in international negotiations for us to say we're not going to do a damn thing and we can continue saying that," he says, "but for our own local domestic reasons let's start doing the right things. In my view, and this might not be popular, we need to reduce the rate of growth. I think it would be terribly wrong for us to emulate what the Western world has done."

His voice is measured, his face shows no sign of impatience. But his restless feet give him away. "There's so much to do and we have very little time," he says. Is he afraid that his own personal time is running out? "I don't know how long I can keep going at this pace," he admits as he hurries back to his car. "I suppose I'm blessed that my energy levels and endurance limits are still high. But I don't know how long that will last."

Pachauri clearly has no intention of slowing down. His next ambition is to mount a major global assessment of energy and the poor: close to 2 billion people still have no access to electricity. "We need to start planning for institutional structures, pricing arrangements, technological initiatives. Otherwise everyone will just sit on their haunches and wait for the electricity line to come their way."

It is easy to miss the energy beneath the calm assertiveness — an energy that, although now tied firmly to 'the cause', surely predates it. Five years ago both his detractors and those supporting him under a misapprehension missed that passion, and the get-things-done focus that it powers. Maybe they should have watched him on the cricket pitch.

"He's a match winner for us," says Tanmay Sharma, a 24-year-old employee at TERI who was surprised at first to find his 67-year-old director-general on the same cricket team. "Just recently we had a match where we required a few wickets urgently and he came into bowl. He's a very deceptive bowler. From the look of it you'd think that his ball is coming very slowly to you and you could just hit him wherever you want to. This was a mistake that the opposition made. They lost all the vital wickets. We won the match." ■

Gabrielle Walker is the author of *The Hot Topic*, with David King, to be published in 2008.

See Editorial, page 1127.

Come all ye scientists, busy and exhausted. O come ye, O come ye, out of the lab

SIR — Will you be working on Christmas Day this year? We have found evidence that increasing numbers of scientists are swapping party hats for mouse mats during the festive season.

We searched for 'Received on 25 December [year]', using Google Scholar, to find out how many submissions were made to academic journals on Christmas Day between 1996 and 2006. Even taking into account the overall increase in the volume of submissions, there were about 600% more manuscripts received by journals on 25 December in 2006 than in 1996 (Figure 1). This trend is still apparent when papers submitted from institutions in Europe and North America are considered in isolation.

We suggest four potential reasons for this move towards seasonal workaholism among scientists.

First, we are collectively falling victim to the 'publish or perish' institutional culture, in which our professional success depends almost exclusively on our publication record.

Second, the pressure on scientists to publish is paralleled by an increase in their administrative and teaching workloads. This pushes research and, in particular, writing into vacation periods.

Third, with the wide-scale implementation of electronic submission systems in the late 1990s, most journals are now 'open for business' every day of the year.

Finally, the demography of religious beliefs among scientists submitting papers may have changed over the years.

Although Christmas Day seems to be an ideal opportunity to get on with some blissfully uninterrupted research, we would urge our fellow scientists to keep their laptops turned off and enjoy a bit of Christmas spirit. You never know, Santa might then be more inclined to bring you that most popular of presents — a paper published in *Nature*!

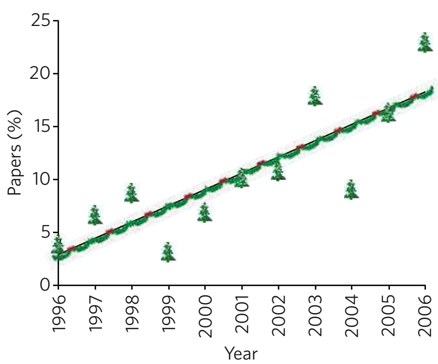


Figure 1 | Can't see the wood for the trees? Take a break. Proportion of published papers submitted on 25 December relative to mean number submitted on the 25th of the month (excluding weekends) for all other months in that calendar year. $R^2 = 0.69$.

Richard J. Ladle†, Ana C. M. Malhado†, Peter A. Todd*
† Oxford University Centre for the Environment, South Parks Road, Oxford OX1 3QY, UK
*Department of Biological Sciences, National University of Singapore, 117543 Singapore

Public control could be a nightmare for researchers

SIR — Last night I had a nightmare. In my dream, all the recommendations made by Pierre-Benoit Joly and Arie Rip in their Essay 'A timely harvest' (*Nature* **450,** 174; 2007) became a reality here in the United States. The public were consulted and actively engaged in practical scientific matters.

I dreamed that the dos and don'ts of science and research were dictated democratically by the American public, of whom 73% believe in miracles, 68% in angels, 61% in the devil and 70% in the survival of the soul after death (see www.harrisinteractive.com/harris_poll/index.asp?PID=618). In my dream, this majority dictated through vigorous 'public engagement' that science should deal with virgin birth, the thermodynamics of hell, the aerodynamics of angel wings, and the physiology and haematology of resurrection.

Suddenly, I found myself in my old lab. There my students were not dealing with the prevalence of gene duplication in bacterial evolution, but were engaged in a heated argument on the virtues of old-Earth versus new-Earth creationism. I woke up in a cold sweat, thinking of what Bishop Samuel Wilberforce's wife reputedly said when confronted with Darwin's theory: "Let us hope it is not true. But if it is, let us hope it does not become widely known."

If Jolie and Rip's proposal for public engagement is workable, let's hope no one ever finds this out.

Dan Graur
Department of Biology and Biochemistry, University of Houston, Houston, Texas 77204-5001, USA

A brief proposal, 150 words

SIR — A common belief about the ways science and art differ is that science convinces with evidence, whereas art persuades through rhetoric. Therefore, we tell ourselves, style does not matter in scientific writing. And yet, of course, it does. Even scientists wish scientists would write more readably.

One place to start is to avoid long words where short ones will do. However, science journals paradoxically foster the use of long words through their 150-word limits on abstracts. Authors who exceed the limit may spend hours looking for bulky portmanteau words to replace several simple ones.

This unfortunate practice has a simple solution. If journals change their word-count limit to a character-count limit that does not include spaces (a function available on most word-processors), scientists will suddenly have an incentive to use short words.

Can such a rhetorical constraint shorten word length? The next call for abstracts could test its effect.

Alice Flaherty
Department of Neurology, Harvard Medical School and Massachusetts General Hospital, Boston, Massachusetts 02114, USA

Tracking early arrivals on the *Nature* trail

SIR — Following the 'History of *Nature*' timeline on your website to 1875, I arrived at the section on 'Continental copycats', which lists European publications with the same name that appeared after the birth of *Nature* in 1869 (www.nature.com/nature/history/timeline_1870s.html). It mentions, among others, a Dutch version; the illustrated monthly *De Natuur* was published from 1881 to 1930.

You don't, however, mention the German journal *Die Natur.* This was founded before *Nature*, in 1852, by teacher–journalist Otto Ule and botanist–pharmacist Karl August Müller. In total, 51 volumes of *Die Natur* were published before it closed in 1902. Issues are held in Kiel University library and elsewhere.

Do readers know of any other titles older than Germany's *Die Natur* and this esteemed journal?

Helmut Zacharias
Windmuehlenberg 6, 24631 Langwedel, Germany

Space travel's no problem, but mice? They're scary

SIR — In the Futures story 'A better mouse-trap' (*Nature* **450,** 456; 2007), a team of scientists, four men and two women, enter an abandoned space station. Upon encountering mice, "one of the women screamed, and the other jumped on a chair" and "three of the men started cursing".

How interesting that even in the distant imagined future of AD 3014, gender stereotypes persist (although, to be fair, one of the men fainted).

Eng H. Lo
20 Staniford Street, Newton, Massachusetts 02466, USA

Contributions to this page may be sent to correspondence@nature.com. We welcome comments on publishing issues at Nautilus http://blogs.nature.com/nautilus).

COMMENTARY

Professor's little helper

The use of cognitive-enhancing drugs by both ill and healthy individuals raises ethical questions that should not be ignored, argue **Barbara Sahakian** and **Sharon Morein-Zamir**.

Today there are several drugs on the market that improve memory, concentration, planning and reduce impulsive behaviour and risky decision-making, and many more are being developed. Doctors already prescribe these drugs to treat cognitive disabilities and improve quality of life for patients with neuropsychiatric disorders and brain injury. The prescription use of such drugs is being extended to other conditions, including shift-workers. Meanwhile, off-label and non-prescription use by the general public is becoming increasingly commonplace.

Although the appeal of pharmaceutical cognitive enhancers — to help one study longer, work more effectively or better manage everyday stresses — is understandable, potential users, both healthy and diseased, must consider the pros and cons of their choices. To enable this, scientists, doctors and policy-makers should provide easy access to information about the advantages and dangers of using cognitive-enhancing drugs and set out clear guidelines for their future use. To trigger broader discussion of these issues we offer the following questions, to which readers can respond in an online forum. Now, on to the questions.

NIKOS/ZEFA/CORBIS

Morning pick-me-up: will drugs that help you stay alert become as widely acceptable as coffee?

Should adults with severe memory and concentration problems from neuropsychiatric disorders be given cognitive-enhancing drugs?

We believe the answer is a resounding yes. A large debilitating aspect of many neuropsychiatric disorders is cognitive impairment. Thus, cognitive-enhancing drugs are a useful therapy option for several disorders, including Alzheimer's disease and Attention Deficit Hyperactivity Disorder (ADHD).

Alzheimer's disease is a neurodegenerative disease of the ageing mind characterized by a decline in cognitive and behavioural functioning, and in particular learning and memory. There are, at present, no treatments for Alzheimer's disease that can stop or reverse the decline in brain function, but cholinesterase inhibitors are being used to ameliorate the impaired neural transmission in the cholinergic system. Such drugs aim to increase the levels of acetylcholine, a neurotransmitter important for maintaining attention and in forming new memories, and may have additional neuro-protective effects.

"The chief concern cautioning against the use of medications is adverse side effects."

Countries with ageing populations are seeing a surge in the number of people with Alzheimer's. The personal and social costs are staggering and in the United Kingdom, economic costs associated with dementias[1] are estimated to rise to £10.9 billion (US$22 billion) by 2031. According to a report commissioned by the Alzheimer's Research Trust in Cambridge, UK, treatment that would reduce severe cognitive impairment in older people by just 1% a year has been estimated to cancel out all predicted increases in long-term care costs due to the ageing population[1].

For all medications, the chief concern cautioning against their use is adverse side effects that affect the individual's health and well being. These may range from mild, temporary physical symptoms, such as dry mouth and headaches, to more severe side effects such as vomiting and joint pain and even cardiac arrhythmia or psychosis. All medications also carry contraindications for certain conditions, such as high blood pressure, when one should not take the drug. For patients with neuropsychiatric disorders, the benefits of the drugs must be weighed against the potential short-term and long-term side effects, and these factors should be discussed with the individual's doctor to ascertain the level of acceptable risk in each case.

If drugs can be shown to have mild side effects, should they be prescribed more widely for other psychiatric disorders?

We believe that cognitive-enhancing drugs with minimal side effects would also benefit many of the patients with schizophrenia, a condition for which they are not yet routinely prescribed. Currently, the disorder affects about 24 million people worldwide.

As with Alzheimer's, the personal and social costs are immense, with economic costs in the United States estimated in the tens of billions of dollars[2]. It is common knowledge that people with schizophrenia typically have hallucinations and delusions, yet it is the long-term cognitive impairments that often impede everyday function and quality of life for many patients. Even small improvements in cognitive functions could help patients with schizophrenia

make the transition to independent living[3].

Thus, cognitive-enhancing drugs are increasingly being considered as possible add-ons to antipsychotic medication, and long-term clinical trials are underway with drugs such as modafinil, which promotes wakefulness[4]. Although the mechanisms of modafinil are not fully understood, it has been found to have direct and indirect effects on various neurotransmitter systems. Behaviourally, an acute dose of modafinil has been found to increase alertness, memory and planning in healthy young adults and cognitive flexibility in patients with chronic schizophrenia[5].

Due to the stated economic and personal costs, the pharmaceutical industry is targeting drugs that would improve impaired cognition in specific neuropsychiatric disorders. Often when a drug is approved for one disorder, its efficacy in improving cognition in additional disorders is investigated and thus its use can be extended to multiple patient groups. In our view, the original justification for drug treatment improving quality of life still holds in these other disorders.

Do the same arguments apply for young children and adolescents with neuropsychiatric disorders, such as those with ADHD?

At present, children diagnosed with ADHD are routinely prescribed long-term medications including atomoxetine and stimulants, such as methylphenidate (Ritalin) and amphetamine. Both methylphenidate and atomoxetine increase the levels of the neurotransmitter noradrenaline. Generally, the therapeutic effects of these drugs include reductions in inattention, hyperactivity and impulsivity, although their widespread and long-term use in younger children has been controversial.

ADHD is a heritable and disabling condition characterized by core cognitive and behavioural symptoms of impulsivity, hyperactivity and/or inattention. ADHD affects 4–10% of children worldwide, and is the most prevalent neuropsychiatric disorder of childhood. ADHD is associated with increased levels of drop-outs from education, job dismissal, criminal activities, substance abuse, other mental illness and accidents[6]. Long-term drug treatment seems to be beneficial in many cases.

However, the side effects of chronic drug use may only become noticeable in the longer term, for example, with apparent reductions in normal growth rates in children with ADHD who are taking stimulant medication[7]. In fact, for many drugs there is limited information on long-term effects and in many areas the findings are inconsistent[7]. Consequently, in all the cases outlined above, we believe the medical consensus would be that medication choice, dose and timing, therapeutic effects and safety should be monitored for individual patients by a healthcare professional. This is particularly important because of potential drug interactions, and so we do not advocate self-medication.

Quick fix: but what are the long-term side effects?

PHOTOTAKE/ALAMY

Would you boost your own brain power?

Cognitive-enhancing drugs are increasingly being used in non-medical situations such as shift work and by active military personnel. This is where the debate about their use begins in earnest. How should the use of cognitive-enhancing drugs be regulated in healthy people? Should their use always be monitored by healthcare professionals?

If offered by a friend or colleague, would you, the reader, take a pill that would help you to better focus, plan or remember? Under what conditions would you feel comfortable taking a pill, and under what conditions would you decline?

"Most would not consider that an espresso confers an unfair advantage at work."

The answers to such questions hinge on many factors, including the exact drug being discussed, its short-term and long-term benefits and risks, and the purpose for which it is used. There are instances in which most people would agree that the use of cognitive-enhancing drugs should be prevented or at least regulated and monitored, such as by healthy children or in competitive settings (including entrance exams to university).

Is it cheating to use cognitive-enhancing drugs?

R. SEAGRAVES/PHOTONICA/GETTY

There are also situations in which many would agree that the use of drugs to improve concentration or planning may be tolerated, if not encouraged, such as by air-traffic controllers, surgeons and nurses who work long shifts. One can even imagine situations where such enhancing-drug-taking would be recommended, such as for airport-security screeners, or by soldiers in active combat. But there are no straightforward answers and any fruitful debate must address each situation in turn.

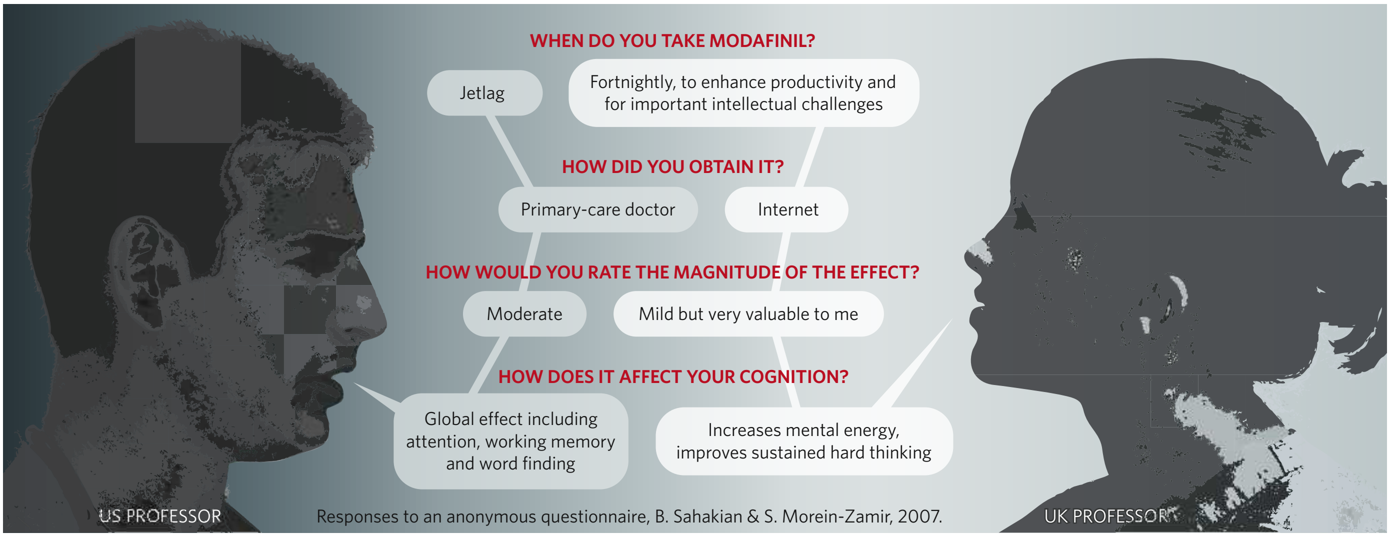
How would you react if you knew your colleagues — or your students — were taking cognitive enhancers?

In academia, we know that a number of our scientific colleagues in the United States and the United Kingdom already use modafinil to counteract the effects of jetlag, to enhance productivity or mental energy, or to deal with demanding and important intellectual challenges (see graphic opposite). Modafinil and other drugs are available online, but their non-prescription and long-term use has not been monitored in healthy individuals.

For many, it seems that the immediate and tangible benefits of taking these drugs are more persuasive than concerns about legal status and adverse effects. There are clear trends suggesting that the use of stimulants such as methylphenidate on college campuses is on the rise, and is becoming more commonplace in ever younger students[8]. Universities may have to decide whether to ban drug use altogether, or to tolerate it in some situations (whether to enable all-night study sessions or to boost alertness during lectures).

The debate over cognitive-enhancing drugs must also consider the expected magnitude of the benefits and weigh them against the risks and side effects of each drug. Most readers would not consider that having a double shot of espresso or a soft drink containing caffeine would confer an unfair advantage at work. The use of caffeine to enhance concentration is commonplace, despite having side effects in at least some individuals[9]. Often overlooked in media reports on cognitive enhancers is the fact that many of the effects in healthy individuals are transient and small-to-moderate in size. Just as one would hardly propose that a strong cup of coffee could be the secret of academic achievement or faster career advancement, the use of such drugs does not necessarily entail cheating.

Cognitive enhancers with small or no side effects but with moderate enhancing effects that alleviate forgetfulness or enable one to focus better on the task at hand during a tiring day at work would be unlikely to meet much objection. And does it matter if it is delivered as a pill or a drink? Would you, the reader, welcome a cognitive enhancer delivered in a beverage that is readily obtainable and affordable, and has a moderate yet noticeable effect



Responses to an anonymous questionnaire, B. Sahakian & S. Morein-Zamir, 2007.

ER PRODUCTIONS/CORBIS

on your concentration and alertness?

How should society react?

When imagining the possible influences of efficient cognitive enhancers on society as a whole, there can be many positive effects. Such drugs may enable individuals to perform better and enjoy more achievements and success. However, cognitive enhancers may have a darker side. Fears have been raised of an overworked 24/7 society pushed to the limits of human endurance, or of direct and indirect coercion into taking such drugs. If other children at school or colleagues at work are taking cognitive-enhancing drugs, will you feel pressure to give them to your children or take them yourself? What if a perfectly safe and reliable cognitive enhancer existed, could society deny it to healthy individuals who may benefit from it?

At present, relatively safe cognitive enhancers with clear effects in healthy individuals are available. Today, in healthy individuals, most cognitive-enhancing drugs yield only moderate effects, and enhance only a subset of cognitive abilities. In the case of some drugs, such as methylphenidate, there are improvement in some domains such as attention, but there may be impairments in others, such as previously learned spatial tasks[10]. Consequently, we believe that current debates must focus on the risks and harms at the level of the individual.

> **"If others at school are taking these drugs, will you feel pressure to give them to your children?"**

In future, drug treatment may be better tailored to individuals through a better understanding of how genes influence the body's response to drugs. Because domain-specific effects vary between individuals depending on their genetic make-up, drug efficacy may ultimately be enhanced and side effects reduced. Many believe that with increasingly sophisticated and targeted treatments, truly smart drugs with moderate-to-large effects on cognition, will become feasible in the future. With the advent of such cognitive enhancers, the discussion must turn to the effects on society as a whole.

There have been debates over the potential of such drugs to decrease or increase disparity in society. Even today many people benefit from their financial status to obtain a better education and nutrition that in turn can enhance brain power. Thus, the potential for cognitive enhancers to increase disparity in society would seem unavoidable.

We believe it would be difficult to stop the spread in use of cognitive enhancers given a global market in pharmaceuticals with increasingly easy online access. The drive for self-enhancement of cognition is likely to be as strong if not stronger than in the realms of 'enhancement' of beauty and sexual function.

Some agreement and regulation on supervised use requires urgent attention, particularly because of concerns in regards to healthy children and adolescents and the long-term consequences of stimulant use on the developing brain. For example, if young stimulant users were informed of the potential effects on their growth, would they continue with long-term use?

Rather than individuals purchasing substances over the Internet, we believe it would be better to ensure supervised access to safe and effective cognitive-enhancing drugs, particularly given potentially dangerous drug–drug interactions. Such regulation must be evidence-based and a product of active dialogue between scientists, doctors, ethicists, policy-makers and, importantly, the general public. This may necessitate a new form of regulation because the remits of the existing health and drug agencies are for the regulation of medicines for treatment, not for enhancement.

Regulations may also be proposed for particular situations in which the use of cognitive-enhancing drugs would impart an unfair advantage, such as in competitive situations, similar to existing regulations for sporting events.

Ultimately, our drug use is a reflection of our society and should never be considered without the broader context of why healthy people choose to use the drugs in the first place. There are other options available for coping with everyday stresses, such as improving the work–life balance or relaxation. Sleep, psychological and behavioural problems can be addressed through alternative methods including psychological treatments, and enhanced cognition can be obtained through education and other means.

Similarly, drug treatment for patients with neuropsychiatric disorders should always be considered as part of a wider treatment programme (including psychological, rehabilitative and social components). Nevertheless, we believe that as we move into the twenty-first century, a key challenge for the pharmaceutical industry will be the development of more effective cognitive-enhancing drugs, so desperately needed by those who have impairments in cognitive and behavioural functioning due to neuropsychiatric disorders or brain injury. ■

Barbara Sahakian and Sharon Morein-Zamir are at the University of Cambridge, Department of Psychiatry and the MRC/Wellcome Trust Behavioural and Clinical Neuroscience Institute, Cambridge CB2 2QQ, UK.

1. Comas-Herrera, A., Wittenberg, R., Pickard, L., Knapp, M. & MRC-CFAS. PSSRU Discussion Paper 1728 (2003).
2. Wu, E. Q. *et al. J. Clin. Psychiatry* **66,** 1122–1129 (2005).
3. Davidson, M. & Keefe, R. S. *Schizophr. Res.* **17,** 123–129 (1995).
4. Morein-Zamir, S., Turner, D. C. & Sahakian, B. J. *Schizophr. Bull.* doi:10.1093/schbul/sbm090 (2007).
5. Turner, D. C. *et al. Neuropsychopharmacol.* **29,** 1363–1373 (2004).
6. Barkley, R. A. *Attention-Deficit Hyperactivity Disorder: A Handbook for Diagnosis and Treatment* 3rd edn (Guilford, New York, 2006).
7. Swanson, J. M. *et al. J. Am. Acad. Child Adolesc. Psychiatry* **46,** 1015–1027 (2007).
8. Farah, M. J. *et al. Nature Rev. Neurosci.* **5,** 421–425 (2004).
9. Lorist, M. M. & Tops, M. *Brain Cogn.* **53,** 82–94 (2003).
10. Elliott, R. *et al. Psychopharmacol.* **131,** 196–206 (1997).

Competing financial interest statement
Barbara Sahakian consults for several pharmaceutical companies and for Cambridge Cognition. She also has shares in CeNeS.

Join the discussion at: http://network.nature.com/forums/naturenewsandopinion/

BOOKS & ARTS

Our social roots

We share many behavioural traits with our primate relatives — some disquietingly nasty.

Gorilla Society: Conflict, Compromise and Cooperation Between the Sexes
by Alexander H. Harcourt and Kelly J. Stewart
University of Chicago Press: 2007. 416 pp. $75 (hbk), $30, £19 (pbk)

Macachiavellian Intelligence: How Rhesus Macaques and Humans Have Conquered the World
by Dario Maestripieri
University of Chicago Press: 2007. 192 pp. $25, £14

Chimpanzee Politics: Power and Sex Among Apes
by Frans de Waal
25th Anniversary Edition. Johns Hopkins University Press: 2007. 276 pp. $24.95, £16.50

G. ELLIS/MINDEN PICTURES/FLPA

Family feast: endangered mountain gorillas (*Gorilla gorilla beringei*) enjoying a vegetarian menu together.

Sarah F. Brosnan

Why do you spend more time with your colleague next door than the one down the hall? As a founding scholar of primate social behaviour, the fifteenth-century philosopher Niccolò Machiavelli might have been able to tell you. Today's primatologists are still fascinated by the evolutionary roots of power, sex and politics in human and non-human primates — surprising parallels emerge that may explain facets of our behaviour and codes governing our society.

A seminal book in the field is Frans de Waal's *Chimpanzee Politics*, just re-released as a 25th-anniversary edition. De Waal explores interactions among three high-ranking males in the Arnhem Zoo colony in the Netherlands to obtain insight into alliances, sex and power in our closest living relatives. The chimpanzees' lives include all the intrigue and shifting allegiances of the Florentine court; it is easy to forget that the participants are not human.

A quarter of a century after its first publication, the influence of de Waal's approach is pervasive. Dario Maestripieri's engaging new book, *Macachiavellian Intelligence*, argues that social cognition is the key to our species' extraordinary success. The book is also a salutary reminder that we are members of the Order Primates as much as of the Family Hominidae, and not all that different from our disquietingly nasty cousins.

Rhesus macaques and humans, Maestripieri explains, are group-living generalists who succeed by advancing their own — and their family's — future through political manoeuvring. Altruism and social behaviour are therefore useful only when the pay-off is greater than the investment, although, according to Maestripieri, humans may have recently evolved more pervasive pro-social tendencies.

Some may question Maestripieri's pragmatic approach to human behaviour, such as his view that our sexual patterns were shaped to secure partner commitment. But, his use of anecdote, and comparisons between humans and macaques, make a persuasive case that a self-interested desire to manipulate others motivates much of human behaviour.

An understanding of how society determines the behaviour of individuals calls for an examination of an outgroup that varies in its degree of relatedness or its social organization. Gorillas, with their harem societies and lesser aggression, provide a nice counterpoint to chimpanzees and macaques (excepting, perhaps, the little studied but apparently more gregarious western gorilla). *Gorilla Society* aims to develop a socio-ecological framework for understanding the animals' social organization and behaviour.

Harcourt and Stewart's book contains some novel approaches. For example, the authors attempt to model rarely seen behavioural variants in order to estimate their pay-offs, which helps in understanding previously unexplained behaviour. They also approach social organization from the male and female perspectives, developing a picture of infinite regress as the decisions of each sex affect each other's choices. They explain, among other things, the conspicuous absence of male takeovers in gorilla populations. Every chapter ends with a comparison between gorilla behaviour and that of chimpanzees, bonobos and orangutans in similar circumstances, illustrating the broader power of socioecological theory.

The authors of all three books are noted primatologists. Although aimed at different audiences, the books are all readable and

informative. There is some repetition in Harcourt and Stewart's because it is written as a reference work; extensive cross-referencing and helpful section headings make it easy to use. Maestripieri's slimmer volume will appeal to a general audience with its fast pace, references to popular culture and wide-ranging discussion of human behaviour. It cites the original studies, but could leave primatologists wishing for more in-depth discussion.

Just as we are on the brink of a more nuanced and thorough understanding of primate and human society, the breakdown of human society continues to fuel the demise of the remaining strongholds of primates in the wild. For instance, gorillas are now listed as critically endangered by the World Conservation Union (*Nature* **449,** 127; 2007).

Contrary to his stereotype, Machiavelli believed that force should be mitigated with prudence, that morality must not be abandoned. Where is our prudence and morality when we ignore the fate of other peoples and species who share our planet? Humans should find a way to narrow the gap between our own well-being and that of our fellow creatures. ■

Sarah F. Brosnan is an assistant professor in the Department of Psychology, Georgia State University, PO Box 5010, Atlanta, Georgia 30302-5010, USA.

Taxonomy on trial

Trying Leviathan: The Nineteenth-Century New York Court Case That Put a Whale on Trial and Challenged the Order of Nature
by D. Graham Burnett
Princeton University Press: 2007. 304 pp. $29.95, £17.95

Henry Nicholls

When the Catholic Church put Galileo on trial for his heretic views, man's position in the Universe was at stake. When schoolteacher John Scopes entered a Tennessee courtroom in 1925 for violating the state's anti-evolution statute, the issue was man's relationship to the animal kingdom. It's hard to imagine that a case brought by a Manhattan fish-oil inspector against a purveyor of whale oil could end up in similar territory. As D. Graham Burnett's enthralling book demonstrates, it did just that.

Trying Leviathan chronicles a courtroom drama that gripped New York in 1818: the case of *Maurice* v. *Judd*. Earlier that year, the state had introduced a law obliging those selling 'fish oils' to have their casks inspected and certified. It was hoped that this would raise the bar for a product often diluted with sediment and dirty water. For James Maurice, New York City's inspector of fish oils, whale oil was part of his remit. Traders such as Samuel Judd, head of the New York Spermaceti Oil and Candle factory, begged to differ and put the law to the test.

The ensuing legal tussle had to settle the superficially simple question of whether whale oil was fish oil and therefore liable for inspection. Burnett, a historian of science at Princeton University, shows it did far more — fuelling a sensational public debate that put the order of nature, and how we know it, in the dock. The impact of this case may, in part, explain why thinkers such as William Whewell and John Stuart Mill applied their minds to precisely this problem of whether a whale is a fish, and why Herman Melville's Ishmael pondered the question in *Moby-Dick*.

Burnett scrutinizes the central whale/fish conundrum from the four main perspectives of the time: laymen, natural philosophers, whalers and businessmen. This conjures up a rich and rounded vision of life in the young republic. He brings New York and the packed courtroom alive, characterizes the state of scientific knowledge at this interesting moment in the history of natural history, cuts deep into cetacean flesh, and steps into the wheeling, dealing commercial world of downtown Manhattan.

Two characters take centre stage in the trial: distinguished naturalist Samuel Latham Mitchill and prosecution lawyer William Sampson. As the star witness for the defence, Mitchill brought the latest science into the courtroom, using linnean taxonomy and state-of-the-art Cuvierian comparative anatomy to make the case for the whale as a mammal. Sampson mugged up on his history of science, and his cross-examination of Mitchill is an awesome display of eloquence and wit. *Trying Leviathan* is worth reading if only for the extraordinary exchange between these men.

Mitchill's faith in comparative anatomy painted an enlightened picture of whales to the jury — viviparous, lactating beasts with complex hearts, lungs and livers. Sampson countered by satirizing the subjective nature of taxonomic division. Who exactly, Burnett asks on Sampson's behalf, had decided that "the intestine preoccupations of comparative anatomy" produced a more accurate taxonomy than one based on visible external characteristics? It's a question as pertinent today as it was in 1818, with the present generation of taxonomists doing their best to assess the value of different types of data, from morphological measurements to genetic sequences.

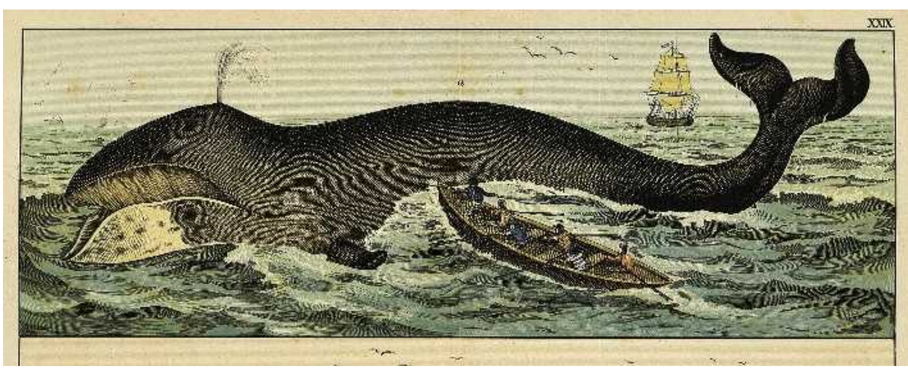
Appropriately, Burnett takes the debate out of the courtroom, out of nineteenth century New York, and into more universal territory. Amid much else, the author questions whether Linnaeus really brought order to the taxonomic chaos as rapidly and thoroughly as is commonly assumed. Zoological classification, Burnett suggests, could still be "divided, contentious, even fickle" decades after Linnaeus.

It's a shame that this nuanced history has come too late to feed into this year's tercentennial of Linnaeus's birth. Thankfully, its vivid portrayal of a pre-darwinian world gives a valuable context for anyone intent on Darwin glorification for his bicentenary in 2009. When Mitchill confidently asserted that "a whale is no more fish than a man", he strayed well into what is normally considered Darwinian territory, bridging the sacred void between the animal kingdom and *Homo sapiens*. It's fascinating to see how this statement (and others like it) played out some 30 years before the publication of *On the Origin of Species*.

Burnett is right to describe *Maurice* v. *Judd* as a "mini-bonanza". Newspapers and periodicals found much in the case to poke fun at. In fact, it would be hard to make this fascinating case a dull read. It's also hard to imagine how anyone could do a better job: he marshals a wealth of primary material — biblical commentaries, school primers, university lecture notes, published taxonomies, sailors' logbooks and private letters.

Burnett curates the abundant quotations with skill and strengthens his thesis with some marvellous contemporary illustrations. His clear writing and delightful detours help build a sense of suspense at the outcome of the trial. All of which makes this serious book an unexpected page-turner. ■

Henry Nicholls is a editor of *Endeavour* magazine and author of *Lonesome George: The Life and Loves of the World's Most Famous Tortoise* (Pan, 2007).



G. DAGLI ORTI/ BIBLIOTHÈQUE DES ARTS DÉCORATIFS PARIS/ THE ART ARCHIVE

Mammal or fish? The question provoked a notorious court case in the early nineteenth century.

Magic squares and circles

Benjamin Franklin's Numbers
by Paul C. Pasles
Princeton University Press: 2007.
266 pp. $26.95, £15.95

Jared Wunsch

Benjamin Franklin was a brilliant amateur scientist in an era when amateur science mattered. His experiments on electricity won him the Royal Society's Copley Medal in 1753 — it is to Franklin that we owe the notion of positive and negative charges. He charted (and named) the Gulf Stream. His prodigious inventions include the lightning rod, the glass harmonica, bifocals and the Franklin stove.

In *Franklin's Numbers*, a book mixing intellectual history and mathematical puzzles (with solutions appended), Paul Pasles brings out a less-celebrated sphere of Franklin's intellect. He makes the case for the founding father as a mathematician. Franklin's main contribution to the subject was his work on magic squares, and these are the focus of the book.

Even bent rows (1, 8, 13, 12, for instance) of Franklin's notes sum to 34.

A magic square is, in its traditional formulation, an $n \times n$ grid containing the numbers 1 to n^2, such that all rows, all columns, and both diagonals sum to the same number. Franklin, characteristically, improved on the usual form, producing squares that could be summed in more intriguing ways, along 'bent rows', for example. He also concocted several magic circles as a further novelty.

Franklin was diffident on the subject of his work on magic squares, sheepishly admitting to having spent time on them out of proportion to the subject's utility. Pasles takes up the defence of Franklin's squares, correctly pointing out that utility is not a suitable measure for a piece of mathematics and that future applications are notoriously hard to predict.

It is here that the case starts to become shaky. The number theorist G. H. Hardy wrote in his *Mathematician's Apology* in 1940 that "the best mathematics is serious as well as beautiful", going on to assert that "the 'seriousness' of a mathematical theorem lies not in its practical consequences ... but in the significance of the mathematical ideas which it connects". By this measure, magic squares, entertaining though they are, rank mathematically just a little higher than chess problems (Hardy's example of real but unimportant mathematics).

Perhaps Franklin just came too late to pure mathematics, already a mature field in his era, but early to electricity, where the work of a gentleman researcher could still be ground-breaking. It was Franklin's electrical work, viewed in the light of Maxwell's equations, that gave us genuine mathematical magic. ■

Jared Wunsch is associate professor of mathematics at Northwestern University, Evanston, Illinois 60208, USA.

Fruits of their labour

Status of Pollinators in North America
by The Committee on the Status of Pollinators in North America, National Research Council of the National Academies
The National Academies Press: 2007.
307 pp. $56

Susan J. Mazer

Forests, prairies, meadows, seashores and wetlands all depend on a diverse and healthy community of pollinators. Wild pollinator populations living in intact and healthy forests, woodlands and fields contribute to the success of a variety of crops — including coffee, watermelons, cucumbers and sunflowers.

It is therefore troubling to hear that pollinator populations are declining or at risk of extinction, the sobering subject of the *Status of Pollinators in North America*. This is a report by a US National Research Council committee, created in 2005 to assess the reality and causes of pollinator declines in agricultural and natural systems, and to offer recommendations to ensure the long-term stability of pollination services.

The committee synthesized the results of some 1,200 research articles and reports, focusing on empirical studies of pollinators and their effects on wild and domesticated plant species. No pollinator escapes: changes in population size and distribution of pollinating ants, bees, wasps, beetles, flies, butterflies, moths, birds and bats are all summarized.

The committee casts an even wider net by enlisting the expertise of applied researchers, honeybee specialists, non-governmental organizations, managers of pollinator databases and industry consultants, who all participated in a symposium at the National Academies in 2005.

A few basic facts highlight the value of domesticated and wild pollinators, and the risks that they face. Honeybees (*Apis mellifera*), the most widely managed, carefully monitored, and commercially distributed pollinator, are used for the fruit and seed production of more than 100 crops (all non-cereals) in the United States. Estimates of their economic value in the United States range from $150 million (at present, the total annual cost of bee-colony rental) to almost $19 billion (the estimated value that farmers would pay if pollinators weren't freely available in nature).

For some crops, honeybees are ineffective pollinators compared with native bees or man-

W. MANNING/CORBIS

Unhappy honeybee: efforts are afoot to restore plummeting pollinator populations.

aged wasps. Some of these alternative pollinators are managed, at a much smaller scale, which itself is risky because rare events — such as disease or environmental change — are more likely to wipe out small populations than large ones. Other domesticated plant species rely exclusively on native bats or birds, whose fate is linked to habitat destruction.

Changes in abundance have been monitored for only a small fraction of the species known to be effective pollinators; there is a growing list of factors implicated in population declines. In the United States, honeybee colonies have more than halved since 1947 (from 5.5 million to 2.4 million). Parasitic mites and pathogens, insecticides used to control crop pests and displacement by Africanized honeybees, are all to blame and may also affect managed populations of non-*Apis* pollinators. Toxic effects of secondary compounds produced by genetically engineered plants are suspected. Habitat modification is probably still the prime culprit in the decline or endangered status of several species of wild pollinators.

A sustained pollinator decline in North America, for example, would mean lower yields from crops that depend on animals for pollination, and so prices would increase; or there would be less variety available as farmers switch from growing insect-pollinated crops to the restricted range of self-fertilizing ones that give reliable fruit or grain production.

Farmers have known for centuries what the public and legislators may be accepting just in time: a field of crops without pollinators is a harbinger of a greater calamity. *Status of Pollinators* offers a host of straightforward and complementary recommendations to help prevent crop failures and the collapse of native plant communities. There isn't a silver bullet to zap the problem: simultaneous application of a variety of solutions will be necessary to sustain a healthy and diverse community of pollinators nationwide.

For example, we need more research entomologists, plant-population biologists, geneticists, agricultural ecologists and systematists. To identify regions vulnerable to pollination failure, we should run international pollinator-monitoring programmes. For long-term pollinator security, Mexico, Canada and the United States should pursue collaborative breeding programmes to identify and manage pollinators other than *Apis*. Land-use practices friendly to pollinators should be adopted by industrial, public and residential landowners. Educational institutions should promote awareness of the intimate connection between plants, their pollinators, our diets and our economy.

In the United States, only a continent-wide commitment to the protection of pollinators will allow future generations to enjoy the fruits of their labour. ■

Susan J. Mazer is professor in the Department of Ecology and Evolution, University of California, Santa Barbara, California 93106, USA.

Glitter bugs

M. LIBBY

Nick Thomas

A pile of dead insects and an assortment of disassembled antique-watch mechanisms would probably be destined for the rubbish heap in most homes. In the pretty coastal town of South Portland, Maine, a young artist combines biology and technology to create bioart sculptures (www.insectlabstudio.com).

Artistic inspiration — he refers to it as his 'epiphany' — first struck Mike Libby in the late 1990s. He happened on a particularly colourful beetle lying dead beside a vending machine. Months later, he assembled his first gear-laden insect from the salvaged workings of an old Mickey Mouse watch, which he transplanted into the beetle.

Libby calls his blending of nature and technology "a celebration of natural and man-made function". He collects local insects such as butterflies, dragonflies and beetles, but many of his specimens now come from companies that supply insect collectors and entomologists, enabling him to adapt his art to exotic species from around the world.

With the eye of an artist and the skill of a surgeon, he replaces the bugs' innards with recycled cogs, springs, dials, steel and brass gears, as well as more modern resistors, capacitors and light-emitting diodes (LEDs). These he carefully glues together to create the bionic bugs. The operation takes 20 to 40 hours.

One of his largest projects, shown here, is the 12-centimetre-long Central American grasshopper, *Tropidacris dux*, more commonly called the giant brown cricket. So large is the wingspan of this mega-insect that hunters have been known to blast it with shotguns, mistaking it for a bird.

Libby adorns *T. dux* with brass and copper parts to complement the grasshopper's orange-brown body and wings. The massive specimen carries a suitably hefty price tag: $950.

With today's microelectronics, a logical extension to Libby's art might be to make it interactive — an LED that actually glows, or a wing that flaps. Yet the artist deliberately avoids applying active electronic components in his works.

"I don't want them to look cheap or toy-like," he says. "Any activity or function should be in the mind and imagination of the viewer." ■

Nick Thomas is an associate professor of chemistry at Auburn University at Montgomery, Montgomery, Alabama 36124, USA.

Multicultural legacy

The Mathematics of Egypt, Mesopotamia, China, India, and Islam: A Source Book
edited by Victor J. Katz
Princeton University Press: 2007. 712 pp. $75.00, £44.95

George Gheverghese Joseph

Mathematics was an important part of many ancient cultures, not just the Greek. These peoples were also at the forefront of notable discoveries, as recorded in *The Mathematics of Egypt, Mesopotamia, China, India, and Islam*.

This pioneering work provides English translations of mathematical texts from each of these regions and cultures, and a better understanding of their contributions to mathematics. There are nuggets of information difficult to find elsewhere. The use of non-mathematical sources, particularly letters and administrative documents from Egypt and Mesopotamia, reveals the practical applications of mathematics and the scribes who composed and used the documents.

Each chapter is devoted to one region and presented by a well-known scholar — Annette Imhausen for Egypt, Eleanor Robson for Mesopotamia, Joseph Dauben for China, Kim Plofker for India and J. Lennart Berggren for the Islamic world. Each author examines the content and impact of primary mathematical sources on the individual cultures. Chapters are roughly structured so that a selection of texts and commentaries follow an introduction, and end with a list of sources and secondary references.

Dauben's study of Chinese mathematics is notable for its length (roughly 200 pages) and

TECHNOLOGY

Earliest known non-lithic tool

As the holiday celebrations get underway, ponder a moment on the small implement bent on prizing an echo of the feast from between your teeth. Henry Petroski was so intrigued by the toothpick that he has written more than 400 pages about its technological and cultural evolution, heavily spiced with historical anecdote (*The Toothpick*, Knopf, 2007).

Chapter titles give the flavour — see 'The tragic heiress'; 'Talking round a toothpick'; 'The fatal martini'; 'The butler did it'.

The toothpick's history stretches back some two million years, as indicated by scored hominid fossil teeth. The Tibetans integrated toothpicks into their jewellery. And in case you ever need to know, New England's Charles Forster invented the first automatic toothpick-making machine in 1869.

Toothpicks over the ages were usually made of metal or wood; goose or porcupine quills were convenient too. Pliny the Younger advised against vulture feathers because "they cause a bad smell", and recommended instead the needle-like bones of a hare.

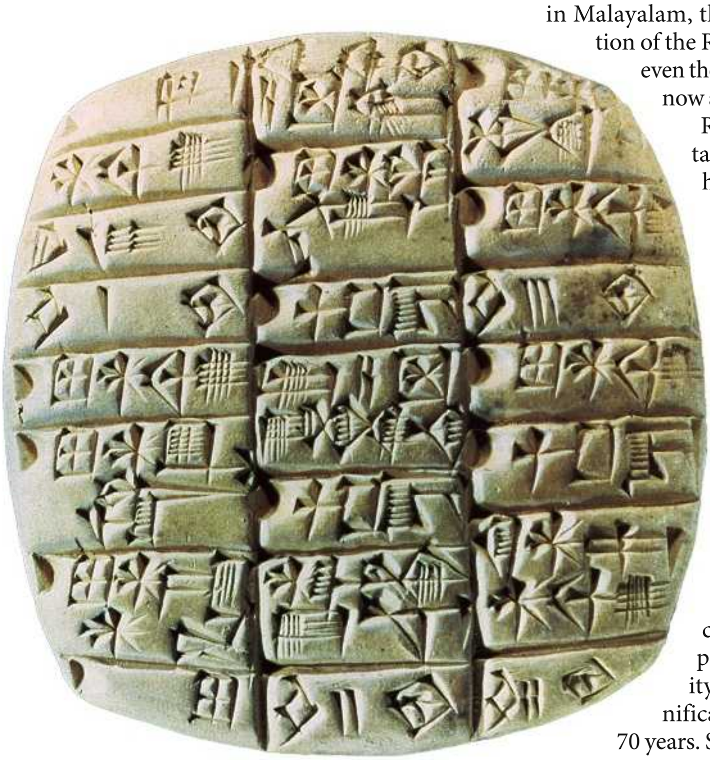
Toothpicking has been victim to both social censure and prescribed etiquette. Pictured here is a Japanese lady from the Meiji era, grasping a toothpick with her fingers held in the accepted pattern.

Happily, the modern toothbrush has barely dented the popularity enjoyed by the toothpick over millennia.

COLLECTION OF H. PETROSKI

clear, comprehensive coverage of the subject. It offers some new and surprising information, such as the recent discovery of the oldest extant text *Suan Shu Shu* (or 'Book of Numbers and Computation'). This dates to around 200 BC and has led to a rethinking of the origins of Chinese mathematics.

Beggren's Islamic mathematics rigorously surveys the sources. It contains an interesting excerpt from al-Kuhi, discussing why and how one should learn mathematics and how to 'prove' results. Both Dauben and Beggren's contributions call into question views expressed by some historians on the role of proofs in non-Western mathematics, particularly in Chinese mathematics.



LOUVRE, PARIS/THE BRIDGEMAN ART LIBRARY

A Mesopotamian accounts table in cuneiform script from 2400 BC.

This brings us to Indian mathematics. Plofker includes Mahavira's (from about 850 AD) paean to the art of calculation in *Ganitasarasangraha* ('Compendium of the Essence of Mathematics'). The period covered is as vast as that of Chinese mathematics, from Vedic Sulbasutras (800–600 BC) to Kerala mathematics (1350–1600 AD), and Plofker uses excerpts from earlier translations as well as from a forthcoming translation of the Kerala text of *Kriyakramakari* ('Operational Techniques') of Sankara Variyar. The problem with fitting such a broad sweep into a single chapter is that excerpts and commentaries have to be tailored, or left out. This is particularly true of Kerala mathematics, where the seminal text in Malayalam, the *Yuktibhasa* ('An Exposition of the Rationale'), is not referred to, even though an English translation is now available.

Robson's chapter on Mesopotamian mathematics contains her own translation of some 60 tablets and an exhaustive discussion of this historical and social context. Example is piled upon example of Babylonian mathematics, some no different from those before. Robson's contribution highlights the issue of what this book is for: is it a conventional source book of translations, or a taster for the curious?

This dichotomy becomes more evident in Imhausen's chapter on Egypt. Imhausen points to the limited availability of sources and lack of significant discoveries over the past 70 years. She makes use of literary and administrative documents that, apart from their mathematical import, provide a useful backdrop to the social and economic concerns of the day. I searched in vain for one of the iconic calculations in Egyptian mathematics (problem 79 in the Rhind papyrus), which is a table listing numbers of houses, cats, mice, wheat and so on, in increasing powers of seven and showing how the sum of this geometrical series can be obtained. Indeed, there are only a few examples of calculation (without an explicit algorithmic basis) and visualization (only two diagrams out of about 15 in the original papyri).

There are a few misprints and verbal infelicities. Maps to accompany each chapter would have been helpful, as would a more comprehensive introduction and an expanded index. All the same, this book is an essential resource for anyone wishing to know more about how the mathematics of the different regions influenced and shaped the development of world mathematics. ■

George Gheverghese Joseph is an honorary reader in the Department of Education, University of Manchester, Manchester M13 9PL, UK. He is author of *The Crest of the Peacock: Non-European Roots of Mathematics* (third edition forthcoming).

New science-art award

Entries are now being accepted for the new Niche Prize, awarded by the Royal Institution of Great Britain and *Nature*. The most original, arresting image or installation conveying a scientific idea in an inspiring, beautiful and artistic way will win the honour of being on prominent display for a year when the Royal Institution reopens next spring following a £20 million (US$41 million) refurbishment. Judges include Susan Greenfield, director of the Royal Institution, Philip Campbell, editor-in-chief of *Nature* and Lady Ritblat. Prizewinners will be part of a key moment in the Royal Institution's 208-year history of celebrating science. The closing date for entries is 22 February 2008. For guidelines and an entry form, see www.rigb.org. To discuss this competition and other science and arts collaborations, see http://network.nature.com/forum/sciart.

The demise of the lone author

As the average number of contributors to individual papers continues to rise, science's credit system is under pressure to evolve.

Mott Greene

Any issue of *Nature* today has nearly the same number of Articles and Letters as one from 1950, but about four times as many authors. The lone author has all but disappeared. In most fields outside mathematics, fewer and fewer people know enough to work and write alone. If they could, and could spare the time and effort to do so, their funding agencies and home institutions would not permit it.

Scientific papers have always contained two quantities — the increment of new science and the credit for its discovery. From the late 1600s until about 1920, the rule was one author per paper: an individual produced an increment of science and obtained a corresponding increment of credit. This symmetry was breached in the 1920s, diminished in the 1950s, and largely abandoned by the 1980s. Collaboration in multidisciplinary research is now universal as well as essential, and we determine from the list of authors who shares in the credit. Curiously, however, in most journals we are not told which of these did what part of the work, nor may we be certain (have we ceased to care?) who drafted the paper.

The ruling convention of multiple authorship is that all authors shared in the work more or less equally and, if the first author or two takes the role of 'first among equals', all listed authors take full credit for the contents of the paper. This is easy enough to swallow where three or four authors are concerned, harder when there are eight to ten authors, and almost impossible with twenty or fifty — let alone hundreds, as in some sequencing papers.

It would, of course, be possible to specify in detail, as in movie credits, who did what on a scientific paper: there is simply no widespread pressure to do so. *Nature*'s editor Philip Campbell introduced a policy in 1999 of including a statement of author contributions in each paper (see *Nature* **399,** 393; 1999). Although this is voluntary, authors in *Nature* are increasingly taking up the option.

Fortunately, where there are large numbers there are laws, and where there are laws there are results to be had. Lotka's law, obtained empirically by the mathematician Alfred Lotka in 1926 and many times confirmed, is a rough 'inverse-square law of scientific productivity'. For every 100 authors who each produce a scientific paper in a given period, there will be 25 authors who produce two, 11 who produce three, and one author who produces ten or more.

The appreciation of Lotka's law has allowed the continuation, in a world of clearly shared credit and hazily specified responsibility, of citation counting as the principal means of establishing scientific prominence and reputation. No matter how many co-authors you have, the more times your name appears on a scientific publication, the more productive you are assumed to be, and the more worthy of support. It can even be shown that Lotka's law predicts the ranking distribution of an author within an author list, and their climb up the scientific ladder.

D. PARKINS

The only natural force opposing the utility of Lotka's law has been Goodhart's law, from the economist Charles Goodhart: "Any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes." Once citation counting became established as a means to determine prominence, players began to 'game the system' based on their knowledge of that standard, and the metric ceased to have a close relation to the outcome it was designed to measure. Such attempts led to the somewhat occult business of impact factors, impact journals, author rank within a paper, and other such countermeasures to re-establish the utility of citation counting.

Until very recently, the combination of Lotka's law and impact factors at least held such 'author gaming' to a draw. Now cracks are appearing in the system. It seems that Lotka's law applies only when papers with 100 or more authors are rare. When these become common, the disjunction between the number of papers being counted and the number of authors entering the system per paper becomes so large that the power-law distribution of multiple authorships breaks down.

The ability of Lotka's law to equate frequency of authorship with scientific rank is buckling as more and more areas of science — genomics, proteomics, climate modelling and particle physics are the most prominent — regularly produce papers with more than 100 authors. Further evolution of the system is likely in the short term.

I predict that in those fields where multiple authorship endangers the author credit system we shall soon see institutionally initiated restriction on the number of authors. Paradoxically, this is likely to be endorsed by all parties as preferable to cinema-style specification of who actually did what. Most will prefer full credit for a few papers to little or no credit for many, considering where it matters most: university committees in charge of tenure, promotion and salary increments based on scholarly production. Given *Nature*'s role in determining, as well as chronicling, how science is reported (see *Nature* **450,** 1; 2007), interested parties could watch these pages to see whether a trend towards more restricted authorship is emerging. ■

Mott Greene is John Magee professor of science and values at the University of Puget Sound, Tacoma, Washington, USA.

This Essay is part of a history website celebrating the launch next month of the complete online archive of *Nature*.
www.nature.com/nature/history

ESSAY

NEWS & VIEWS

PHYSICS

A quantum less quirky

Seth Lloyd

What physicists want for Christmas is a solution to the philosophical conundrums of quantum mechanics. They will be disappointed, but work that dissolves one aspect of quantum weirdness is some consolation.

There is a strange quantum-mechanical phenomenon. A brilliant researcher receives a Nobel prize for work on quantum physics, but then expresses scepticism about the validity of the theory. Albert Einstein started the trend in 1921, and remains its most famous exponent. He received his Nobel for the quantum-mechanical explanation of how electric current can be generated just by shining light on a surface (the photoelectric effect), but for the rest of his life expressed a deep-seated distrust of the quantum. God, as he famously said, does not play dice. Most recently, Tony Leggett, a Nobel laureate in 2003, expressed his belief that quantum physics is at the very least incomplete, and needs to be supplemented by some new physics. Writing in *Physical Review A*, Wojciech Zurek[1] now provides some balsam for these and other pained minds.

So what makes geniuses deny the source of their inspiration, and muddy the font of their fame? The answer is the intrinsic bizarreness of quantum mechanics. Waves consist of particles; particles are shadowed by waves; electrons (or anything else, for that matter) can be in two places at once.

Tests of infant cognition show that the idea that an object cannot be in two places at once is ingrained in our psyches from the age of about three months. At the same age, babies become aware that objects exist even when they cannot be seen. Playing peek-a-boo with a child aged less than three months is intensely dissatisfying: when you cover your face, they exhibit no excitement or interest. Daddy is gone: so what? After three months, everything changes. When you cover your face, the child waits with eager anticipation for the "Boo!": he or she knows you're there behind the hands.

In quantum mechanics, if you can't see an object, you mustn't assume it is there: an unmonitored electron can be, and generally is, everywhere at once. By the age of three months, children are better equipped to live in the macroscopic world, but their intuition for quantum mechanics is spoilt.

Everyone finds quantum mechanics counterintuitive, Nobel laureates included (at least, Nobel laureates older than three months). I find quantum mechanics counterintuitive. I am not a Nobel laureate, but I spend my research time thinking about quantum mechanics in order to build quantum computers and quantum communication systems. But so what if I find it counterintuitive? My intuition is frequently wrong anyway. As long as I can perform the calculations and get the right answers, then I should be happy.

But if ever a scientist deserved to trust his intuition, it was Einstein. For his sake, and for those like him who find quantum weirdness deeply distressing, we should delve a little further into the roots of the problem. In his paper[1], Zurek does just that, providing a simple, elegant and intuitive explanation of one of the strangest and most counterintuitive features of quantum mechanics — its peculiar mathematical shape.

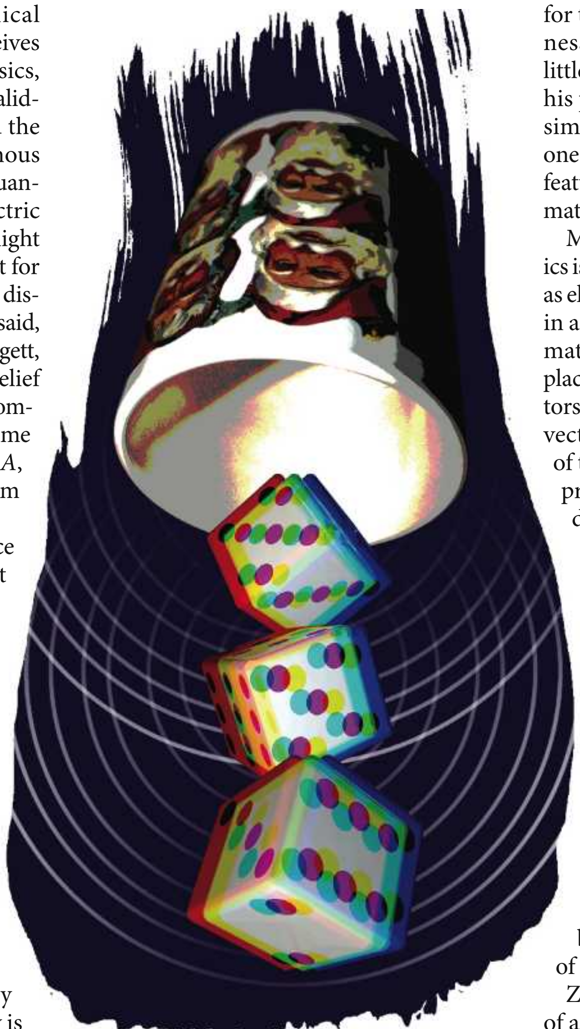


ILLUSTRATION BY ANDY MARTIN

Mathematically speaking, quantum mechanics is a strange beast. The states of particles such as electrons correspond to functions or vectors in a complex vector space. Physical transformations (an electron hopping from place to place, for example) correspond to linear operators or matrices acting on those functions and vectors. The act of measuring the properties of the system equates to applying the appropriate mathematical operator to the vector describing the system. On measurement, the system 'collapses' to an eigenvector of the measurement operator (that is, a vector that the operator, applied again, will simply transform into a multiple of itself), and an associated eigenvalue. The eigenvalue gives the outcome of the measurement.

Why is this so? Why does measurement leave a system described by nothing other than an eigenvector ('in an eigenstate', in the jargon) of the measurement operator? For the past 80 years, the answer to this question has been because Erwin Schrödinger and Werner Heisenberg said so: it's just one of the 'postulates' of quantum mechanics.

Zurek shows that it is in fact a consequence of an intuitively more satisfactory postulate: that if one makes a measurement twice in rapid succession, one always obtains the same result. He uses an argument based on his and William Wootters's proof[2] of the celebrated 'no-cloning' theorem (essentially, that you can't create identical copies of an unknown quantum state) to show that if a measurement were to leave a system in anything other than an eigenstate of that measurement, immediate repetition of the measurement would have a chance of yielding a different result. A quirky, mathematical postulate of quantum mechanics is thus replaced by a simple derivation from an intuitive result.

Einstein was celebrated for proclaiming that God is subtle, but not malicious. The proof that, despite His predilection for games of chance, God does not attempt to change the rules mid-game certainly would have delighted Einstein. Would it have convinced him of the validity of quantum mechanics? My intuition tells me not. ■

Seth Lloyd is in the Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA.
e-mail: slloyd@mit.edu

1. Zurek, W. H. *Phys. Rev. A* **76,** 052110 (2007).
2. Wootters, W. K. & Zurek, W. H. *Nature* **299,** 802–803 (1982).

CANCER DIAGNOSTICS

One-stop shop

Jonathan W. Uhr

Detecting cancer early and monitoring its progress non-invasively are high on oncologists' agenda. So the design of a neat device that detects and counts cancer cells shed into the blood by tumours is a welcome advance.

A hallmark of carcinomas — malignant tumours of epithelial tissues such as breast and colon — is the invasion of neighbouring tissues at an early stage of tumour growth. This process is associated with the shedding of tumour cells into the bloodstream. So counting and characterizing such circulating tumour cells to monitor the progress of the disease might be the most viable alternative to the unacceptable process of repeated, invasive biopsies. On page 1235 of this issue, Nagrath *et al.*[1] report the development of an efficient and selective method for isolating these rare cells using microchip technology.

It might be considered too late to diagnose and characterize carcinomas if tumour cells have already entered the circulation, but this is not always the case. For example, although the presence of circulating tumour cells in patients with breast cancer usually indicates a poor prognosis, it does not necessarily indicate cancer metastasis. This is because only a small fraction of these cells have metastatic potential[2], and soon after entering the bloodstream, many tumour cells become committed to a suicide programme[3]. Also, circulating tumour cells are found in a significant proportion of breast-cancer patients long after such patients have undergone a mastectomy, despite being clinically disease-free and at a very low risk of their cancer recurring[4].

Identifying and counting these cells can also be useful for monitoring carcinomas. For example, patients who have more than a certain level of circulating tumour cells have a poorer prognosis and so require more aggressive therapy[5]. Counting the cells after a course of treatment has been started can show whether it is effective, thereby allowing the treatment regimen to be modified much earlier than if physicians had to rely solely on changes in tumour size[5]. Furthermore, analysis of particular genes or proteins in tumour cells has revealed that they are continually altering genetically. So genetic changes in themselves could be a reason for modifying a patient's course of treatment[6], and could be useful for selecting drugs that target particular cellular pathways involved in the malignancy. Indeed, such patient-specific approaches are deemed the therapies of the future.

The usual criteria for identifying circulating tumour cells are their microscopic appearance (they are larger than most white blood cells, with a large nucleus and little cytoplasm), a characteristic expression profile of the cytokeratin protein, and the absence of a protein marker for normal blood cells. Although not definitive, these criteria are probably sufficient identification if there are large numbers of tumour cells in the patient's blood.

Various techniques have been used to capture these rare cells, and advances include the development of an automated instrument for counting them known as CellSearch, which is currently in clinical use[7]. The detection unit of this instrument uses metal particles coated with an antibody to the EpCAM protein, which is present on virtually all carcinoma cells[8]. A magnetic field is then used to capture antibody-bound tumour cells. However, such techniques generally detect circulating tumour cells in only 20–60% of patients with metastatic disease[9].

The work of Nagrath *et al.*[1] is the latest advance. The authors have developed a microfluidic ('lab-on-a-chip') device that can separate circulating tumour cells from whole-blood samples. The surface of this chip houses 78,000 microposts, each coated with antibodies to EpCAM. Fresh, unprocessed blood is pumped across the chip under controlled flow conditions to avoid damaging the fragile tumour cells. (This is a great improvement on conventional immunomagnetic enrichment procedures, which often damage the cells.) The cells bind to the microposts and are identified by a camera that can detect their morphology, viability and the presence of tumour markers on their surface (see Fig. 1 on page 1235). Different tumour markers can be identified by staining the cells with specific antibody–dye conjugates.

The chip seems to be highly sensitive. The authors could detect circulating tumour cells in almost all of the patients they examined who had recurrent carcinomas, regardless of the tumour's organ of origin, and in all of seven patients with early-stage tumours. Moreover, the purity of the cells obtained with this device was far higher


DR KESSEL & DR KARDON/TISSUES & ORGANS/GETTY

Red and white blood cells in an arteriole. The chip designed by Nagrath *et al.*[1] can readily detect tumour cells among these.

than can be achieved with other methods. But Nagrath and colleagues' one-step method does have its limitations. For example, it cannot examine individual cells for genetic changes associated with progress of the tumour.

Nonetheless, Nagrath and colleagues' results[1] bring us closer to having a fully automated instrument that can detect circulating tumour cells with exquisite sensitivity. Such an instrument would allow routine monitoring of blood for tumour cells as part of a medical examination, and could result in early detection and treatment, with the chance of obtaining higher cure rates. But this approach would also necessitate minimizing false-positive results, particularly when only a few putative tumour cells are isolated. One way to solve this problem would be to count sufficient numbers of chromosomes or genes in candidate cells to detect one of the hallmarks of a malignant cell — a significant deviation from the two-copies-per-cell norm. A simpler alternative might be to test for increased expression of the many known malignancy-associated proteins. In time, it might even be possible to detect and therapeutically target or remove the cells' organ of origin at an early enough stage to prevent the cancer from metastasizing. ■

Jonathan W. Uhr is in the University of Texas Southwestern Medical Center, 6000 Harry Hines Boulevard, Dallas, Texas 75390-8576, USA.
e-mail: jonathan.uhr@utsouthwestern.edu

1. Nagrath, S. *et al. Nature* **450,** 1235–1239 (2007).
2. Fidler, I. J. *J. Natl Cancer Inst.* **45,** 773–782 (1970).
3. Méhes, G. *et al. Am. J. Pathol.* **159,** 17–20 (2001).
4. Meng, S. *et al. Clin. Cancer Res.* **10,** 8152–8162 (2004).
5. Riethdorf, S. *et al. Clin. Cancer Res.* **13,** 920–928 (2007).
6. Meng, S. *et al. Proc. Natl Acad. Sci. USA* **101,** 9393–9398 (2004).
7. Cristofanilli, M. *et al. N. Engl. J. Med.* **351,** 781–791 (2004).
8. Went, P. T. H. *et al. Hum. Pathol.* **35,** 122–128 (2004).
9. Allard, W. J. *et al. Clin. Cancer Res.* **10,** 6897–6904 (2004).

PLANETARY SCIENCE

A younger Moon

Alan Brandon

The most recent study of lunar rocks indicates that the Moon formed later than previously thought — a conclusion that requires our view of the early history of the inner Solar System to be revised.

It was almost 40 years ago that we found out what the Moon is made of: lunar samples collected during the Apollo missions turned out to consist of rocky material similar in composition to that found on Earth. We also think we know that the Moon formed from the accretion of molten and vaporized ejecta produced by a collision between proto-Earth and a giant Mars-sized impactor[1,2]. But the answer to the 'When?' of Moon formation has remained elusive. On page 1206 of this issue, Touboul *et al.*[3] examine this question with tungsten-isotope measurements of lunar rocks.

The mass of the giant impactor that led to the formation of the Moon was as much as 30% of Earth's mass today, and this collision is thought to represent the last significant growth stage of our planet[1]. The post-collision disk of ejecta and the new bulk Earth were a combination of proto-Earth material and that of the impactor. The tremendous release of energy probably melted the Earth, producing a global-scale magma ocean. Iron-rich metal from the proto-Earth and the impactor merged to form the present Earth's core, with Earth's mantle and crust forming from silicate and oxide (non-metal) material. The time intervals between the various events — the giant impact, the formation of Earth's core, and the onset and solidification of magma oceans in Earth and the Moon — tell us when accretion was completed and constrain the early differentiation and cooling histories of the 'terrestrial' planets (Mercury, Venus, the Earth–Moon system and Mars).

Enter the evidence from the ^{182}Hf (hafnium) and ^{182}W (tungsten) radiometric clock. The isotope ^{182}Hf decays to ^{182}W with a half-life of 9 million years. This clock dates events that occurred in the first 60 million years or so of Solar System history, before most ^{182}Hf decayed away. It is useful for tracking metal–silicate differentiation times, because W has an affinity for iron-rich metal cores whereas Hf favours mantle silicate and oxide minerals. If core formation occurred when ^{182}Hf was still actively decaying, then a record is preserved in the amount of ^{182}W present.

The amount of ^{182}W measured in Earth samples[4,5] indicates that W was removed into Earth's core about 30 million years after the Solar System began to accrete 4.567 billion years ago (Fig. 1). Before Touboul and colleagues' new study[3], the lunar ^{182}W data had yielded two conclusions. First, that formation of the Moon — and so the giant impact event — occurred around the same time as Earth's core formed at about 30 million years[4,5]. Second, that the lunar magma ocean largely solidified 10 million years after that[6]. This chronology provides a well-ordered Earth–Moon system that behaved exactly as we expected — rapid accretion of Earth, core formation and global-scale melting were coincident with a giant impact, and were followed by rapid cooling and consequent solidification of the magma oceans of Earth and the Moon.

But some observations do not fit this simple lunar history (Fig. 1). First, radiometric ages for the lunar crust, which was made from light minerals that solidified and floated to the top of the magma ocean, show that it began to form no earlier than 70 million to 150 million years after accretion of the Solar System began[7]. Second, the ^{146}Sm (samarium) and ^{142}Nd (neodymium) radiometric clock shows that the lunar magma ocean largely began to solidify as late as 215 million years[8]. Third, as the new lunar data of Touboul *et al.*[3] show, the previous W isotope data for lunar materials were not completely corrected for ^{182}W produced by the decay of ^{182}Ta (tantalum). The isotope ^{182}Ta forms when cosmic rays bombard the Moon. The excess ^{182}W from this process can result in spurious ages using the ^{182}Hf–^{182}W clock.

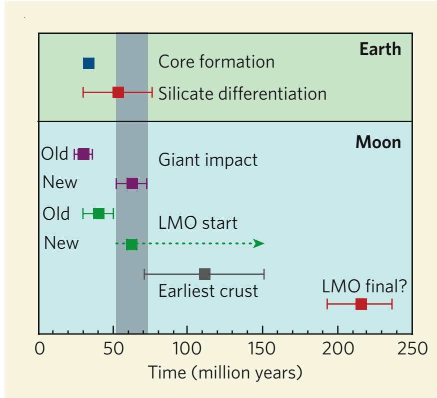


Figure 1 | The earliest history of the Moon and Earth. Time 0 marks the onset of Solar System accretion at 4.567 billion years ago, with the rest of the scale denoting time after that event. The previously inferred ages for the giant impact and the onset of solidification of a lunar magma ocean (LMO)[4–6] were based on older W-isotope data for Moon rocks. Touboul and colleagues' W-isotope data[3] provide a new and later time for the giant impact and for the start of solidification of the LMO. The final solidification time of the LMO is obtained from the ^{146}Sm–^{142}Nd chronology of Moon rocks[8]. The vertical band is the earliest time for the giant impact that formed the Moon; error bars represent 2σ uncertainty estimates for each event. Data points for events on Earth are given for comparison.

When correctly accounting for ^{182}W produced from ^{182}Ta, Touboul *et al.*[3] find that the amount of ^{182}W in lunar samples indicates that the Moon could not have formed before 62^{+90}_{-10} million years after the initiation of Solar System accretion, at least 16 million years after Earth's core formed and no later than formation of the first lunar crust (Fig. 1). These new findings are in agreement with the slower rate of magma-ocean solidification obtained from ages for the lunar crust and ^{146}Sm–^{142}Nd data.

50 & 100 YEARS AGO

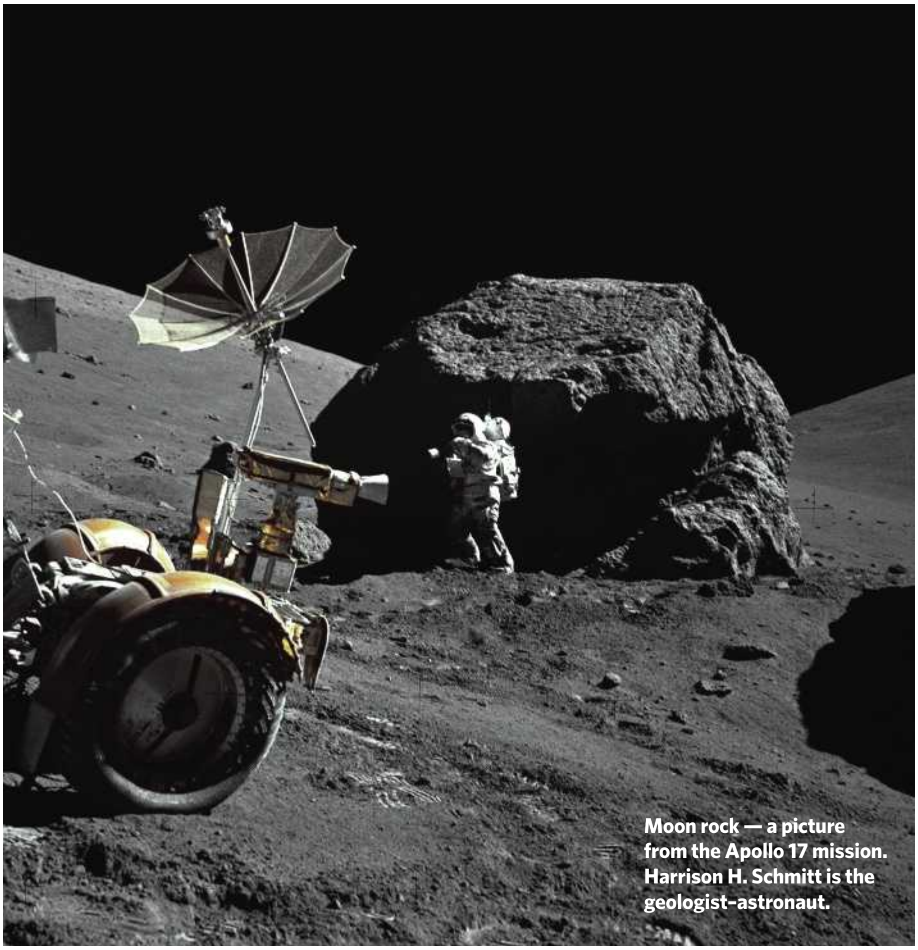
50 YEARS AGO

"An experiment on 'telepathy' using television" by Donald Michie & D. J. West — We have made a small-scale trial with the object of testing any generalized extra-sensory perception effect and looking for individuals with strongly manifested telepathic abilities ... The viewers were informed that the cards would be drawn at random from a pack consisting of three types, depicting a canoe, a wheelbarrow and a trumpet, respectively. They were asked to record their guesses on a form printed in the TV Times and post it to the programme contractors ... The pooled results showed no significant deviation from chance expectation. But one entry, submitted by a Mr. B. Downey, with 15 guesses out of 19 was considered sufficiently suggestive to justify further tests.
From *Nature* 21 December 1957.

100 YEARS AGO

On the day of going to press, we learn of the death of Lord Kelvin, an announcement which will be received with deep sorrow throughout the civilised world ... For the body of one who has brought such honour to the British nation, the only appropriate place of burial is Westminster Abbey.

ALSO:
The increase in the efficiency of colleges and universities in this country is too pressing a need to be dependent upon party politics. Unless our statesmen can be made to realise the supreme importance of this matter and be persuaded to deal with it in a patriotic manner, generously and expeditiously, as if there were no votes to retain or secure, we must reconcile ourselves to the idea that as manufacturing and distributing people we shall in due course have to occupy a third or fourth place among the nations of the world. In Germany, the United States, and now in Japan, rulers have learnt the lesson that efficient education and industrial success are related to each other as cause and effect.
From *Nature* 19 December 1907.



E. A. CERNAN/NASA

Moon rock — a picture from the Apollo 17 mission. Harrison H. Schmitt is the geologist-astronaut.

The later time for the Moon's formation challenges the current view that the terrestrial planets grew rapidly, and also challenges ideas about their early cooling histories[4–6,9]. It may mean that Earth and Mars took at least 50 million years, and possibly hundreds of millions of years, to reach their final mass (that is, 99% of their present size). By contrast, recent models call for a more rapid accretion that took 30 million years or less[9]. Interestingly, the first evidence from the ^{146}Sm–^{142}Nd clock of silicate differentiation in Earth is about 40 million to 60 million years after accretion of the Solar System began[10]. This probably records the time of the onset of cooling of Earth's magma ocean, which overlaps with the earliest new time for lunar formation and the giant impact (Fig. 1). In turn, this implies that earlier core formation in Earth did not necessarily coincide with the development of a magma ocean, and that, for a large terrestrial planet such as Earth, rapid pulses of accretion, which could have been caused by a single giant impactor that ultimately led to the formation of the Moon, might have been necessary to initiate large-scale melting.

The long-lived magma ocean in the Moon raises the question of how long Earth's magma ocean took to solidify. We do not know for certain, but the implication of Touboul and colleagues' lunar W-isotope data[3] is that accretion and early cooling of Earth were not as rapid as we thought. Application of the ^{146}Sm–^{142}Nd clock to martian meteorites suggests that the magma ocean in Mars took 60 million to 100 million years to solidify[11], likewise implying that Earth's magma ocean probably took longer to solidify than some current models predict. We need additional evidence to further examine these issues on the earliest history of the Earth–Moon system. The clues might lie in future samples returned from the Moon and Mars. ■

Alan Brandon is at Astromaterials Research and Exploration Sciences, Johnson Space Center, NASA, 2101 Nasa Parkway, Houston, Texas 77058, USA.
e-mail: alan.d.brandon@nasa.gov

1. Canup, R. M. *Icarus* **168,** 433–456 (2004).
2. Pahlevan, K. & Stevenson, D. J. *Earth Planet. Sci. Lett.* **262,** 438–449 (2007).
3. Touboul, M., Kleine, T., Bourdon, B., Palme, H. & Wieler, R. *Nature* **450,** 1206–1209 (2007).
4. Kleine, T., Münker, C., Mezger, K. & Palme, H. *Nature* **418,** 952–955 (2002).
5. Yin, Q. *et al. Nature* **418,** 949–952 (2002).
6. Kleine, T., Palme, H., Mezger, K. & Halliday, A. N. *Science* **310,** 1671–1674 (2005).
7. Norman, M. D., Borg, L. E., Nyquist, L. E. & Bogard, D. D. *Meteor. Planet. Sci.* **38,** 645–661 (2003).
8. Rankenburg, K., Brandon, A. D. & Neal, C. R. *Science* **312,** 1369–1372 (2006).
9. Jacobsen, S. B. *Science* **300,** 1513–1514 (2003).
10. Bennett, V. E., Brandon, A. D., Heiss, J. & Nutman, A. *Geochim. Cosmochim. Acta* **71,** A79 (2007).
11. Debaille, V., Brandon, A. D., Yin, Q. Z. & Jacobsen, B. *Nature* **450,** 525–528 (2007).

STRUCTURAL BIOLOGY

The dance of domains

Joseph D. Puglisi

Static crystal structures provide only clues about the way large biological molecules work. A recently developed spectroscopic technique also reveals details of their molecular motion, as shown for an RNA molecule from HIV.

The full impact of a ballet can never be conveyed by paintings or photographs — the dynamism of a performance is lost if you can't see the graceful movements of the dancers. The same is true of structural biology, which seeks to reveal the beauty of biological systems at the atomic scale. Observing molecular movement is crucial for us to understand how proteins and nucleic acids work. In this issue, Al-Hashimi and colleagues[1] report a nuclear magnetic resonance (NMR) technique that allows them to see how sections of an RNA molecule move in relation to each other (see page 1263). This approach should provide fresh insight into the structural dynamics of biological molecules.

The mainstay technique of structural biology is X-ray crystallography, which has provided molecular models for the structure of DNA, simple proteins and enzymes, and even for larger assemblies. More recently, other methods — such as electron microscopy at cryogenic temperatures, or NMR spectroscopy — have evolved to provide molecular views of large biological molecules. Each of these techniques places experimental constraints on atomic positions from which a three-dimensional structure is derived. These structural models are immensely powerful, allowing the formulation of theories about substrate binding, enzyme catalysis and regulation in atomic detail. Yet the structural images are like movie stills, shown statically on the page; biophysicists pine for an animated version.

Biological molecules are dynamic machines that change shape to execute their functions. Proteins and nucleic acids are often built from modules known as domains, which move relative to each other when the parent molecule performs a task. Snapshots of domain movements have been captured by 'freezing' molecules into distinct states, usually by binding them to a partner known as a ligand. A mechanistic 'movie' of how the molecule might work is pieced together from the snapshots, like an arcade attraction from the early years of the twentieth century.

To observe molecular motion directly, biophysicists use spectroscopic methods that usually probe the time dependence of light's interactions with part of the molecule. This works because the dynamics of a biological system often modulate the response of spectroscopic probes to light perturbation. For example, NMR spectroscopy is a powerful method for exploring molecular dynamics — the individual nuclei of the atoms that make up the molecule provide rich spectroscopic data.

NMR studies of molecular dynamics have traditionally monitored fluctuations of individual nuclei (or pairs of nearby nuclei) that occur within nanoseconds, which is roughly the timescale of molecular tumbling about an internal axis in solution. More recent experiments[2] have probed dynamics on slower timescales (microseconds to milliseconds) — dynamics that may be linked to the functional states of biomolecules. But these spectroscopic methods focus on the local jittering of atoms, rather than the global movements of large groups of atoms, such as structural domains.

Al-Hashimi and his co-workers[1] now present a technique that observes domain movements directly, on the basis of NMR measurements called residual dipolar couplings (RDCs). Nuclei such as ^{1}H, ^{13}C, ^{15}N and ^{31}P behave like small compass needles when placed in a magnetic field, aligning with the field. Such aligned nuclei can subsequently interact with each other through dipolar interactions. The strength of the interaction depends on the distance between the nuclei and on their orientation with the external magnetic field. In solution, random tumbling of molecules averages the dipolar coupling to zero. But if the tumbling of a molecule is partially restricted such that there is a preferred overall orientation, a small degree of coupling — an RDC — remains, and can be easily measured[3].

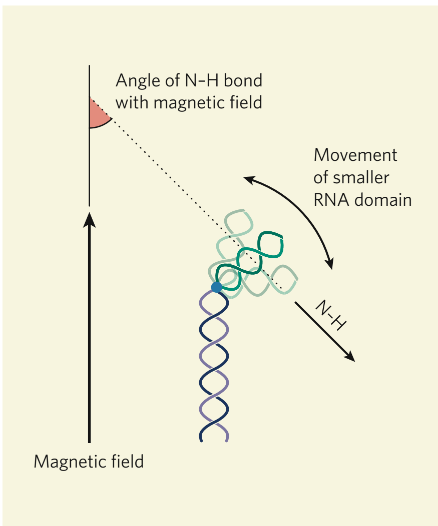


Figure 1 | Molecular alignment and domain motions in RNA. With the help of external agents, RNA molecules in solution will align with a magnetic field. If the molecules consist of two helical domains linked by a hinge, the longer helix dominates the alignment. The magnetic field causes the atomic nuclei in the molecule to behave like tiny bar magnets, and couple to each other through dipolar interactions. By measuring the strength of these interactions, the angle that a chemical bond (here, a nitrogen–hydrogen bond from one of the base pairs in the RNA) makes with the external magnetic field can be measured. Al-Hashimi and colleagues[1] use this effect to show that two helical domains in an RNA molecule move, adopting many different angles to each other.

By using RDCs, the myopia is removed from NMR spectroscopy. Whereas other NMR measurements are based purely on the interactions of nuclei with their immediate surroundings, RDCs provide a broader perspective. The coupling depends strongly on two factors: the distance between two nuclei and (in simple terms) the orientation of the pair of nuclei with respect to the external magnetic field. Because the distance between nuclei is fixed by the geometry of chemical bonds, the variation of RDCs depends only on orientation. For a rigid nucleic-acid helix, with its axis aligned in the direction of a magnetic field, the RDC associated with equivalent pairs of atoms in all the base pairs should be roughly the same. But now imagine that the helix is bent into a boomerang shape, with two rigid sections joined together at a fixed angle. The two sections are no longer aligned in the same orientation to the magnetic field, and the RDCs in the two sections are different (Fig. 1). The use of RDCs for determining the structures of proteins and RNA in solution is now routine, and has certainly improved the quality of NMR-derived structures[4].

Al-Hashimi and colleagues[1] harness RDCs to study HIV TAR — an RNA molecule that consists of two helical domains separated by a hinge of three nucleotides (usually referred to as the bulge). TAR is the binding site for an HIV protein known as Tat; it also binds many ligands at the region near the bulge. Previous structural studies[5,6] have revealed that TAR adopts different conformations depending on which ligands are bound. Furthermore, the structure of unbound TAR has been shown by NMR to be in constant flux. Al-Hashimi *et al.* probe the range of conformations adopted by TAR using RDCs.

The authors had to overcome a common problem that arises when measuring RDCs — the movement of domains not only alters the relative orientations of those domains, but can also affect the overall alignment of the molecule with the magnetic field. Al-Hashimi and colleagues[1] therefore greatly extended one of the two RNA helices in TAR. The elongated helix dominates the alignment of the molecule with respect to the magnetic field, so that the molecular orientation does not change as the second domain moves about. By clever control

of the nucleotide composition, the elongated helix is invisible to the NMR experiment[7].

The authors probed different portions of the TAR RNA by measuring several RDCs in two elongated molecules. Generally, they observed that RDCs in shorter helices were smaller than those in elongated ones. This is consistent with the presence of rapid movements in the shorter helices, because RDCs are damped by molecular motions that occur at sub-millisecond timescales. The damping was reduced when TAR was bound to a ligand that fixes the bulge conformation, which restricts the movements of the RNA domains.

By analysing their RDC measurements using a model of rigid helix movements about the bulge, Al-Hashimi and colleagues mapped the conformational states adopted by TAR over a millisecond timescale. The authors' results clearly show that the two helices do not simply bend at the bulge; instead, they undergo correlated bending and twisting. Strikingly, the conformations of unbound TAR (that is, TAR in the absence of ligands) determined from the RDC data agree well with the conformations observed in crystal structures of TAR in complex with different ligands[8]. This suggests that unbound TAR has evolved to adopt conformations that are favourable for binding ligands, and that these conformations are low-energy states for the molecule.

The methods of Al-Hashimi *et al.*[1] are applicable to several other modular RNA systems; the ability to explore the relative movements of helices in these systems will shed light on RNA folding. The RDC approach can also be used to study protein-domain movements, which will no doubt correlate with protein functions. Using a combination of RDC measurements and complementary fluorescence techniques, it should also be possible to probe domain dynamics over a wider range of timescales than those described in the current paper[1]. And, looking further ahead, Al-Hashimi and colleagues' method will eventually lead to new RDC approaches that do not require the molecules' helices to be elongated. It seems that the production of molecular movies has only just begun. ■

Joseph D. Puglisi is in the Department of Structural Biology, Stanford University School of Medicine, Stanford, California 94305-5126, USA.
e-mail: puglisi@stanford.edu

1. Zhang, Q., Stelzer, A. C., Fisher, C. K. & Al-Hashimi, H. M. *Nature* **450,** 1263–1267 (2007).
2. Palmer, A. G. & Massi, F. *Chem. Rev.* **106,** 1700–1719 (2006).
3. Tjandra, N. & Bax, A. *Science* **278,** 1111–1114 (1997).
4. Lipsitz, R. S. & Tjandra, N. *Annu. Rev. Biophys. Biomol. Struct.* **33,** 387–413 (2004).
5. Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D. & Williamson, J. R. *Science* **257,** 76–80 (1992).
6. Aboul-ela, F., Karn, J. & Varani, G. *J. Mol. Biol.* **253,** 313–332 (1995).
7. Zhang, Q., Sun, X., Watt, E. D. & Al-Hashimi, H. M. *Science* **311,** 653–656 (2006).
8. Ippolito, J. A. & Steitz, T. A. *Proc. Natl Acad. Sci. USA* **95,** 9819–9824 (1998).

PHOTONICS

Charge of the light brigade

David J. Nesbitt

The latest study of fluorescence in nanowires shows that it can be controlled by electric fields. This finding suggests the presence of mobile charge carriers, which could be useful for designing nanoelectronic devices.

There is an irony in the fact that the world of nanostructured materials has become one of the biggest areas of chemistry and physics[1–4], with myriad applications emerging in catalysis, biophysics and the health sciences. A crucial aspect of this is the development of molecular photonic and electronic devices, which are constructed from nanoscale components. But at such tiny sizes, quantum effects have an unpredictable impact on light emission and the flow of electrons (or other charge carriers) in semiconductors. Reporting in the *Journal of the American Chemical Society*, Kuno and colleagues[5] describe how the emission of light from semiconductor nanowires can be controlled by external electric fields, suggesting the presence of mobile charge carriers. This discovery could have implications for the design of 'wires' in nanocircuits.

One of the most promising and intensely investigated areas of nanotechnology is that of inorganic semiconductors, as these materials are small enough to influence the optical and electronic properties of the resulting 'quantum-confined' structures. Nanoparticles of these materials are especially useful because they can efficiently absorb and emit light (fluoresce). As the nanoparticles get smaller, the quantum wavefunctions associated with charge carriers in the structure become compressed. This compression increases the frequencies of light absorbed and emitted by the particles, so that nanoparticles can be designed to fluoresce at specific colours by controlling their size.

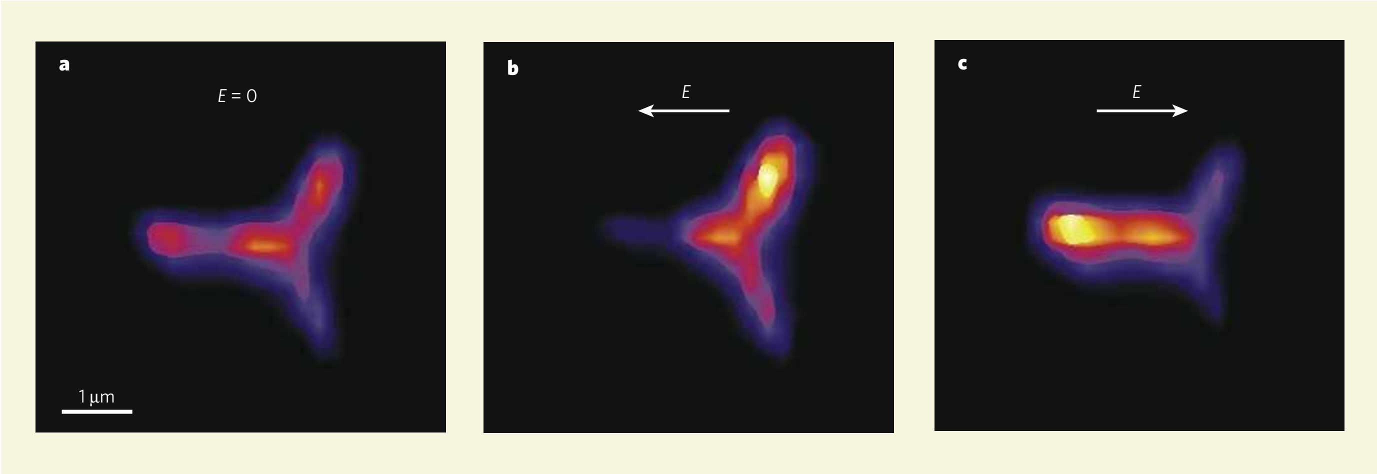
Semiconductor nanoparticles come in many forms — from zero-dimensional quantum dots to one-dimensional quantum wires and two-dimensional quantum wells, with a dizzyingly large inventory of chemical and physical substructures in each category. Quantum nanowires of cadmium selenide (CdSe) are particularly intriguing, as they can be made long enough (1–10 micrometres) for their fluorescence to be imaged using a conventional microscope, but thin enough (5–12 nanometres) to achieve quantum-confinement effects. As such, they are good test systems for exploring the influence of external fields on electric conduction, charging and emission in one-dimensional nanoscale materials.

Kuno and colleagues[5] examined such nanowires in their study. They prepared CdSe nanostructures using a reaction previously used to prepare quantum dots, but they used 'seed' particles of bismuth-coated gold to induce nanowire formation. Several structures could be generated in which one or more wires extended from a seed particle. The authors controlled the topologies of the structures by varying the reaction temperature and the ratio of the starting materials, producing branched shapes such as tripods, V-shapes and Y-shapes.

When the authors subjected the nanowires to an electric field (up to 20 kilovolts per centimetre, a relatively modest value compared with earlier studies of quantum dots[6]), they found to their surprise that the intensity of the fluorescence was enhanced in some regions of the wire and diminished in others — usually two or threefold, but in some nanowires as much as tenfold (Fig. 1). Furthermore, they observed a clear directionality to the effect: light emission was enhanced at the end of the nanowire closest to the positive electrode (where negative charge would accumulate) and diminished at the end nearest to the negative electrode. In the presence of an alternating electric field, the enhancement oscillated by tracking the direction of the field. Although the authors' camera-based detection methods were too slow to directly monitor oscillations with frequencies above 10 hertz, a net fluorescence enhancement due to the applied electric field seemed to be maintained up to frequencies of at least 1 megahertz.

Kuno and colleagues[5] speculate that the fluorescence enhancement is caused by mobile charges that flow rapidly along the nanowire in response to the electric field. They propose that a few of these charges become trapped in defect sites on the surface of the nanowire. Light is emitted from nanowires when excited electrons lose energy and return to the ground state, but non-radiative pathways also exist that are thought to be associated with surface defects. By filling the defect sites, trapped mobile charges could block these non-radiative pathways, forcing energy to be released by light emission.

Although the nanowires show a sustained boost in fluorescence under alternating-voltage conditions, in a stable electric field the enhancement fades away after a few seconds. The authors tentatively attribute this decay to the slow movement of counterions that have opposite charge to that of the mobile carriers, and that can travel along a layer of water molecules that sheaths the nanowire. When the counterions eventually catch up with the faster-moving mobile charges on the nanowire surface, they can balance out the build-up of



AM. CHEM. SOC.

Figure 1 | Enhanced fluorescence in a nanowire tripod. Nanowires made of semiconductors fluoresce. Kuno and colleagues[5] have made branched structures that contain nanowires, such as this tripod made from the semiconductor cadmium selenide. **a**, This image shows the baseline fluorescence of the tripod in the absence of an electric field. **b**,**c**, When an external electric field is applied, enhanced fluorescence occurs in regions of the tripod that are closest to the positive electrode; fluorescence is diminished at the opposite side. The authors suggest that mobile charge carriers in the semiconductor cause this effect.

charge that presumably causes the enhanced fluorescence, thereby quenching it.

With their working hypothesis in mind, Kuno and colleagues[5] went on to study the effect of electric fields on the motion of nanowire bundles floating in solvents. The authors found not only that the bundles rotated into alignment with the applied fields, but also that their centres of mass moved towards either one of the electrodes. This behaviour can be explained only if mobile charges are present in the nanowires, rather than static dipoles — in agreement with the authors' theory. Furthermore, the authors were able to estimate the net surface charge per nanowire by modelling the various factors associated with the movement, such as torques, rotational speeds and solvent viscosities. They also showed that, surprisingly, the sign of the net charge changes depending on the solvent: the nanowires accumulated excess negative charge in some solvents, but positive charge in others.

Kuno and colleagues' work might provide fresh insight into a long-standing question about semiconductor nanoparticles: why does the fluorescence of CdSe quantum dots blink on and off[3]? The model usually invoked to explain such blinking relies on the idea of electron–hole pairs. It is useful to imagine that, when an electron in a semiconductor is promoted to an excited state by a photon, a positively charged 'hole' also forms, which behaves like a mobile charge carrier. In a 'switched-on' quantum dot, fluorescence occurs efficiently when electrons and holes recombine. But sometimes an electron (or a hole) can become trapped on or near the surface of the dot[7,8]. The corresponding hole (or electron) in the core of the dot can then lose energy in a rapid non-radiative pathway that competes with fluorescence, so that the dot 'switches off' and becomes dark. The charge carrier on the surface can eventually return to the core, restoring the natural uncharged state, and thus the on–off cycle of fluorescence can repeat itself extensively.

Tantalizing support for this model has been provided by sensitive experiments[9] showing that the charge on quantum dots 'blinks' on and off in a way that is clearly analogous to the observed intermittent fluorescence. But it has proven surprisingly difficult to link these two phenomena unambiguously. By providing direct evidence for the correlation between electric field, fluorescence and the net accumulation of surface charge on nanoparticles, Kuno and colleagues' results[5] might provide a crucial step towards explaining the root causes of these mysterious blinking phenomena. ■

David J. Nesbitt is at JILA, National Institute of Standards and Technology, and the Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309, USA.
e-mail: djn@jila.colorado.edu

1. Murray, C. B., Norris, D. J. & Bawendi, M. G. *J. Am. Chem. Soc.* **115,** 8706–8715 (1993).
2. Alivisatos, A. P. *Science* **271,** 933–937 (1996).
3. Nirmal, M. *et al. Nature* **383,** 802–804 (1996).
4. Peng, X. *et al. Nature* **404,** 59–61 (2000).
5. Protasenko, V., Gordeyev, S. & Kuno, M. *J. Am. Chem. Soc.* **129,** 13160–13171 (2007).
6. Empedocles, S. A. & Bawendi, M. G. *Science* **278,** 2114–2117 (1997).
7. Efros, A. L. & Rosen, M. *Phys. Rev. Lett.* **78,** 1110–1113 (1997).
8. Verberk, R., van Oijen, A. M. & Orrit, M. *Phys. Rev. B* **66,** 233202 (2002).
9. Krauss, T. D. & Brus, L. E. *Phys. Rev. Lett.* **83,** 4840–4843 (1999).

NEUROSCIENCE

Neighbourly synapses

Bernardo L. Sabatini

Experiences shape our behaviour, memories and perception. Mechanistically, they also influence the brain's circuitry, and cooperativity between neuronal contacts during learning may contribute to this process.

Neuronal plasticity describes experience-related and development-associated structural and functional changes in the brain, which contribute to, among other processes, memory formation. Such changes occur at many levels; for example, depriving an animal of visual stimuli results in both small-scale modifications in neuronal receptors, and large-scale rewiring of neural circuits[1,2]. Many of these changes involve plasticity at the level of synapses, the specialized contact points between neurons, and much of the tremendous learning power of the mammalian brain is thought to arise directly from the vast number of synapses that it contains. So understanding the degree to which each synapse can be independently regulated is necessary to comprehend the computational capacity of our brains. Harvey and Svoboda[3] (page 1195 of this issue) carried out micrometre-scale analysis to determine whether the induction of an activity-dependent form of synaptic plasticity, known as long-term potentiation (LTP), at one synapse alters the properties of its immediate neighbours.

Neurons communicate with each other through neurotransmitters. A stimulated neuron (the presynaptic neuron) fires an action potential, which results in the release of neurotransmitters from its axon into the synaptic cleft between it and another neuron (the postsynaptic neuron). Released neurotransmitters bind to receptors on the postsynaptic neuron

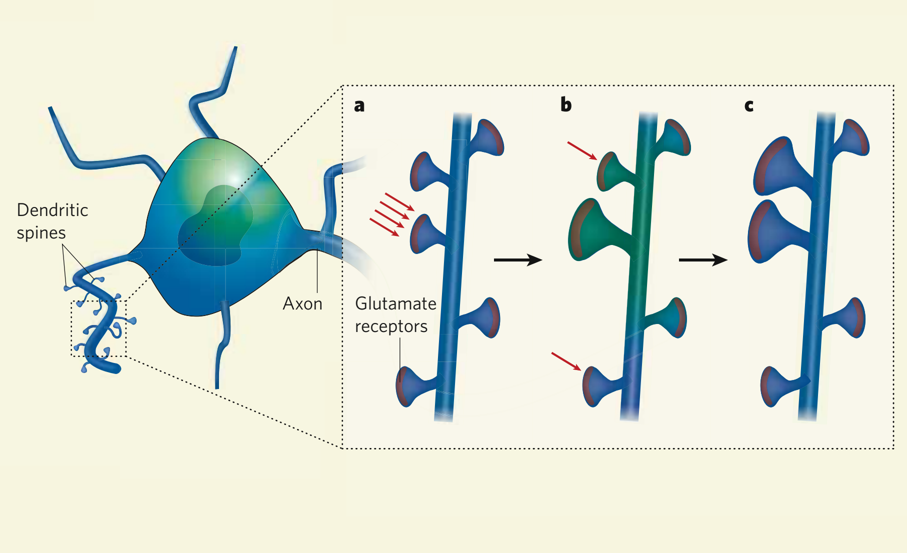


Figure 1 | Cooperativity between synapses. Five dendritic spines, each containing the same number of postsynaptic glutamate receptors, are shown. **a**, When a strong stimulus (red arrows) induces long-term potentiation (LTP) at a single spine, it causes the stimulated spine to grow (**b**). Harvey and Svoboda[3] find that LTP induction at one spine also lowers the threshold for LTP induction at its immediate neighbours (within a distance of 10 micrometres). Consequently, weak stimulation of a nearby spine (red arrow) is sufficient to induce LTP and functional and structural plasticity in the neighbouring synapses (**c**). Weak stimulation of the lower spine that lies outside this zone, however, does not induce plasticity. Approximately 10 minutes after the original stimulation, the plasticity threshold returns to normal and spine growth is stabilized.

and allow ion entry. Excitatory synapses are often made onto dendritic spines (small projections along dendrites) on the postsynaptic neuron. LTP describes strengthening of synaptic communication between two neurons, observed when they are persistently activated simultaneously.

Studying synapses in the hippocampus of the mouse brain, Harvey and Svoboda focused on the effects of LTP on synapses that are located close to the potentiated synapse but that are not stimulated directly. Similar work has been done previously, but the earlier findings[4–6] have been controversial because, depending on sample preparation and the protocol for LTP induction, synapses were reported to be depressed, unaffected or potentiated.

The problem with most previous studies on the effects of LTP on unstimulated synapses is that the experimenter did not choose which synapses to study, and instead analysed those that happened to be stimulated by an electrode. Axons in the hippocampus are divergent and meandering, so even weak stimuli delivered through an electrode placed close to the cell of interest stimulate synapses dispersed throughout large sections of the dendritic arbour. Consequently, the spatial arrangement and the number of actively communicating synapses remain unknown, making it difficult to probe the independence of synaptic plasticity at a fine scale.

But technological advances allow delivery of stimuli to postsynaptic termini in various spatio-temporal patterns. The trick is to bypass the presynaptic terminal and use light to trigger a chemical reaction that produces the neurotransmitter glutamate and directly stimulates the dendritic spine. Addition of a large, light-absorbing side group to glutamate makes it inactive, or what is known as caged; light pulses can be used to rapidly break down the cage, releasing the active neurotransmitter. For example, two-photon excitation is used to trigger glutamate uncaging and release in sub-femtolitre (less than 10^{-15} litres) extracellular volumes. Thus, glutamate is delivered to individual spines located relatively deep within the brain tissue, under conditions that mimic the time course of synaptically released glutamate.

This approach was previously used[4] to study synaptic plasticity in slices of brain tissue maintained in culture and led to the discovery that LTP induction at a single postsynaptic terminus is accompanied by stable growth of the stimulated spine. The structural and functional plasticities were limited to the stimulated spine, indicating that both types of plasticity are synapse-specific.

Harvey and Svoboda[3] asked whether LTP induction at one spine might alter the ability of neighbouring spines to undergo plasticity. They find that, after LTP induction, although structural and functional plasticity is limited to the stimulated spine, surprisingly, the threshold for the induction of plasticity in its immediate neighbour synapses is dramatically reduced. This effect is seen in spines located within roughly 10 micrometres of the stimulated spine and lasts for about 5 minutes (Fig. 1). Furthermore, the phenomenon is robust, because it occurred not only in glutamate-uncaging experiments but also in response to electrical induction of LTP, which is more similar to the physiological setting.

The mechanism underlying the reduced threshold for plasticity in neighbouring synapses is unknown. It could be simple, with LTP induction at one spine activating a pathway that enhances synaptically evoked calcium influx into neighbouring spines; alternatively, it may involve more complex adjustments of the LTP-induction machinery. But one point is certain — spatially clustered groups of synapses act in a mutually reinforcing and cooperative way.

These results[3] may also help to explain the findings of a previous study[5], which showed that potentiation triggered by strong synaptic stimulation spreads to non-stimulated synapses. In hindsight, instead of potentiation spreading, it is probable that the threshold for the induction of plasticity in the adjacent synapses was lowered; consequently, subthreshold stimulation, which is normally used to probe the state of neighbouring synapses, was enough to induce LTP, giving the impression that potentiation had spread.

Harvey and Svoboda's findings open several avenues of research. Computational and theoretical neuroscientists will probably find that lowering of the LTP threshold at neighbouring synapses, in response to LTP of one synapse, endows neural networks with new behavioural and learning capabilities[7]. Cell biologists will be fascinated by the mechanisms underlying the 'priming' of neighbouring synapses for LTP induction. For example, they would want to know what molecules are released from the potentiated spine and what pathways they activate on arrival at the neighbouring spines. Electrophysiologists would want to know whether the spatial and temporal spread of the effect on the LTP threshold depends on the number of synapses potentiated. Furthermore, they would be curious to establish how the effects described by Harvey and Svoboda relate to synaptic capture and tagging. (These are mechanistically distinct phenomena that occur over long timescales and in response to strong stimulation of many synapses, and allow the potentiation of one set of synapses to alter the threshold for plasticity induction at others.)

Returning to the question of the organizational scale of regulation in the brain, these findings[3] indicate that synapses located close to each other are neighbourly, acting, to some extent, as a cooperative unit. Co-activation of neighbouring synapses on a single dendritic segment is known to trigger local action potentials that act as spatially delimited, associative signals for the induction of LTP[8]. Owing to these mechanisms, synapses that tend to fire together and that are near each other on a dendrite will be potentiated more readily than simultaneously active synapses separated by large distances. If potentiation increases the likelihood that a synapse is preserved during development, this may lead to

TECHNOLOGY

The art of illumination

In November 2006, the Yamaguchi Prefectural Art Museum in southern Japan held an exhibition entitled 'The Trip to Sesshu'. It commemorated the 500th anniversary of the death of the Zen Buddhist monk and painter Sesshu Toyo, whose delicate *suibokuga* ink paintings have been designated 'National Treasures' by the Japanese government.

So that visitors could appreciate Sesshu's scrolls as closely as possible to the way original admirers did — under candlelight or torchlight — some of the paintings were illuminated by a specially designed light-emitting diode (LED) system. Tsunemasa Taguchi and Michitaka Kono now provide the technical details of that system (T. Taguchi and M. Kono *J. Light Visual Environ.* **31,** 149–151; 2007).

LEDs are solid-state light emitters known for their energy efficiency, flexibility of design and robustness. For a long time, they were made to emit light only in a particular part of the visible-wavelength spectrum. But in the mid-1990s a new generation emerged, based on blue LEDs covered with a yellow phosphorescent layer, which emitted bright, white light.

Taguchi and Kono's lighting system used special LEDs that contained three different phosphors, each emitting at different frequencies. This meant that the white light

emitted had a particularly high colour quality, as quantified on a scale known as the colour-rendering index. To optimize viewers' appreciation, the authors tailored the LEDs to render the earthy red colours characteristic of the antique scrolls especially well.

The individual lights were positioned so as to distribute light evenly on the artwork (first two paintings pictured), rather than scattering it around them as fluorescent lamps would do (paintings in background). The output of the LEDs was stable and did not cause heating, thus assisting preservation of the precious scripts.

White LEDs are becoming ever brighter, more efficient and less expensive. As a result, traditional light bulbs are increasingly on the way out. Large-scale applications, such as car lights, traffic signals and Christmas decorations, are where the economic benefits of LED use are being felt. But the illumination of antique Japanese art must surely rank as one of the diodes' most aesthetically pleasing applications.

Liesbeth Venema

T. TAGUCHI & M. KONO

the spatial clustering of synapses with similar firing patterns. Thus, for some forms of activity-dependent plasticity in the hippocampus, the fundamental unit of regulation might be larger than an individual synapse[9], and rather a physically clustered cohort of synapses with similar firing patterns, whose spatial arrangement on dendrites arises naturally following mutually reinforcing interactions between synapses.

It is neuroscientists' goal to understand the plastic features of the brain that make storing memories and learning new behaviours possible. In trying to achieve this formidable goal, many neuroscientists hope that, by uncovering the mechanisms behind the regulation of individual synapses, they will reveal the rules that govern the wiring of the brain. Harvey and Svoboda's results introduce a new level of complexity. They demonstrate that these rules vary across distances as short as a few micrometres, and are themselves altered on a minute-by-minute time frame. ■

Bernardo L. Sabatini is in the Department of Neurobiology, Harvard Medical School, 220 Longwood Avenue, Boston, Massachusetts 02115, USA.
e-mail: bsabatini@hms.harvard.edu

1. Shatz, C. J. & Stryker, M. P. *J. Physiol. (Lond.)* **281,** 267–283 (1978).
2. Heynen, A. J. *et al. Nature Neurosci.* **6,** 854–862 (2003).
3. Harvey, C. D. & Svoboda, K. *Nature* **450,** 1195–1200 (2007).
4. Matsuzaki, M., Honkura, N., Ellis-Davies, G. C. R. & Kasai, H. *Nature* **429,** 761–766 (2004).
5. Engert, F. & Bonhoeffer, T. *Nature* **388,** 279–284 (1997).
6. Abraham, W. C. & Goddard, G. V. *Nature* **305,** 717–719 (1983).
7. Poirazi, P. & Mel, B. W. *Neuron* **29,** 779–796 (2001).
8. Golding, N. L., Staff, N. P. & Spruston, N. *Nature* **418,** 326–331 (2002).
9. Govindarajan, A., Kelleher, R. J. & Tonegawa, S. *Nature Rev. Neurosci.* **7,** 575–583 (2006).

METROLOGY

New generation of combs

Steven T. Cundiff

To measure an optical frequency, you are best off using an optical frequency comb. A radical approach shakes up how these combs are produced, and will permit their closer integration into optical-fibre technology.

Increasing the range over which frequencies can be accurately measured is an exertion driven both by applications (making sure that a mobile phone uses the right channel, for instance) and by fundamental physics: time and frequency are the most accurately measured physical quantities, and thus are often used in tests of theories such as relativity and quantum mechanics. Optical frequency combs[1,2] — arrays of regularly spaced, well-defined reference frequencies — have revolutionized these endeavours[3,4]. Optical atomic clocks using this form of benchmarking can be more precise[5] than the very best clocks referenced to the current caesium atomic time standard, and combs are increasingly being used for sensitive and rapid detection of molecular processes[6]. The growing importance of the technique was recognized by its appearance in the citation for the 2005 Nobel Prize in Physics[7].

On page 1214 of this issue, Del'Haye *et al.*[8] describe the creation of an optical frequency comb using a toroidal glass microresonator. The very strong light fields produced drive the optical response of the glass into the nonlinear regime, where the principle of wave superposition no longer applies and waves can mix with one another to create new frequencies. The flexibility and small size of this apparatus give it huge potential for use in diverse areas, from telecommunications to astrophysics.

Frequency combs provide a way of accurately measuring frequencies that are too high to be measured directly. At radio frequencies, combs are produced by driving an electrical element, typically a step-recovery diode, with a simple, sinusoidal input signal. Once this sine wave exceeds a threshold, the diode converts it into a square wave, which contains new frequencies. The highest frequency is determined by the switching time, and is around 100 picoseconds for the best diodes. The standardized comb frequencies thus generated are integer multiples (harmonics) of the well-known input-wave frequency.

Generating a comb with a useful frequency spacing at optical frequencies requires a different approach. Until a few years ago, this meant injecting light from a continuous-wave laser into an optical cavity containing an electro-optic modulator driven by a radio-frequency signal[9]. The result was a cascade of evenly spaced frequency lines above and below the laser's optical frequency, corresponding to adding or subtracting integer multiples of the radio-frequency modulator signal, with each line generating the next. Combs generated in this way generally spanned a frequency range of several terahertz.

NATURE|Vol 450|20/27 December 2007

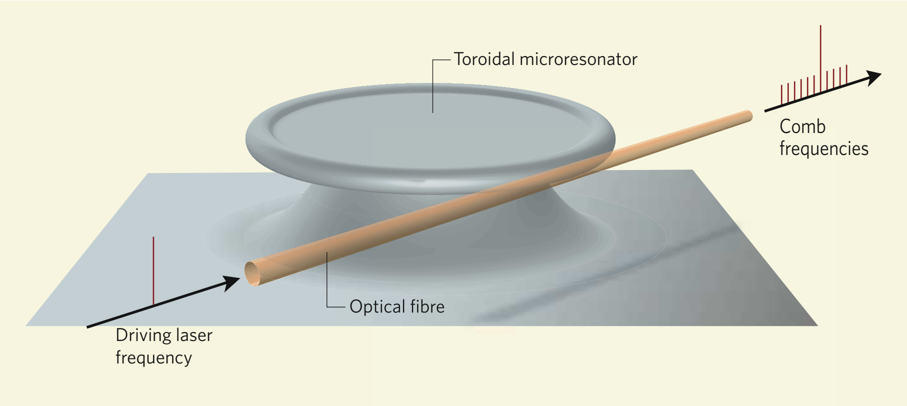


Figure 1 | Unfurling the standard. The high 'quality factor' of the toroidal microresonator used by Del'Haye *et al.*[8] allows a driving laser field to be held for a long time, greatly increasing the light intensity to the point at which the response of the resonator's glass becomes nonlinear. When a continuous wave laser, corresponding to a single, well-known frequency spike (left), is launched into this microresonator, its light is converted into a regularly spaced comb of frequencies (right) through its interaction with the wave modes of the resonator cavity. Such combs allow frequencies to be measured accurately over a wide range of frequencies (in this case optical), as each 'tooth' acts as a reliable frequency standard.

The philosophy changed in the year 2000, with the advent of femtosecond (10^{-15} s) optical frequency combs[1,2]. These combs are produced by mode-locked lasers whose light comes in short, sharp bursts of femtosecond duration. A train of such short pulses can be decomposed into light at the laser's offset frequency (the rate at which its phase evolves) plus integer multiples of the laser's repetition rate (the number of pulses produced per second). The shorter the pulse, the wider the frequency spectrum spanned by these frequency lines. Under the right conditions, a spectrum spanning hundreds of terahertz — broad enough to cover the entire visible spectrum, and thus look white to the eye — can be produced.

The breadth of the band is crucial, because most methods for measuring the offset frequency use a self-referencing technique that requires a spectrum spanning an octave (a factor of two in frequency)[10]. The great advantage of the femtosecond-laser approach is that it locks the phases of all comb lines together. By contrast, the electro-optic modulator locks the phases only of adjacent comb lines, allowing phase variations to build up towards the edges of the comb, which limits the precision of any frequency measurement there.

In some ways, the new approach to optical-frequency generation taken by Del'Haye *et al.*[8] is similar to older radio-frequency techniques. Like the radio-frequency approach, it takes a sinusoidal input signal — the output of a continuous-wave laser at a wavelength of 1,550 nanometres — and couples it to a nonlinear response medium, in the form of a toroidal microresonator[11]. This microresonator stores up the laser light, greatly increasing its intensity as more and more light enters such that nonlinear wave mixing occurs. The frequencies of the output lines from the microresonator are equally spaced on either side of the frequency of the input laser light, with a spacing determined by the properties of the microresonator (Fig. 1).

The nonlinear generation of a pair of frequency lines either side of an optical signal is not new. The generation of tens of them is, and Del'Haye *et al.* achieve it by using a microresonator with a very large 'quality factor' — a measure of the sharpness of its resonant response. This not only enhances the resonator's nonlinearity but also produces a cascading effect, with each frequency line generating the next, just as in the electro-optic comb.

For a spectral comb to be useful as a frequency reference, the spacing of the comb lines must be perfectly even, and the lines themselves must be narrow. Conservation of energy and momentum show why the spacing of lines in a toroidal microresonator must be even: in the initial nonlinear response, two 'pump' photons of frequency ν_p are destroyed to generate one photon in each of the sidebands. These have frequencies ν_+ and ν_-, which are related by $2\nu_p = \nu_+ + \nu_-$, according to energy conservation. This allows a continuum of frequencies; momentum conservation, which expresses itself as the requirement that the phase evolution of the two lines be matched, then selects out one pair of frequencies for the first two lines. Once the initial pair of lines is generated, the cascade that then builds up preserves their spacing.

By comparing the comb generated by a microresonator to a femtosecond comb, Del'Haye and colleagues demonstrate that the spacing is indeed regular, to better than one part in 10^{-17}, rivalling the quality of femtosecond combs[12]. The implication, although not yet definitely proved, is that the comb lines are narrow, and thus phase noise is not building up significantly in the nonlinear cascade.

An obvious advantage of the new device is its small size and potential for integration with optical-fibre technology. But the offset frequency of the comb being determined by the frequency of the pump laser is a double-edged sword: if the laser frequency is known and stable, it is an advantage; if it is unstable, it is a disadvantage. If the frequency of the pump laser is not known, it can be determined using the same self-referencing technique used to determine the offset frequency in a femtosecond comb. This technique is easiest to apply if the width of the comb is pushed to cover an octave (in Del'Haye and colleagues' set-up[8] it is already close).

Probably the biggest obstacle to using the authors' comb for optical-frequency metrology is the spacing of its teeth, which is around 1 terahertz. The ideal comb spacing is of the order of the detection bandwidth, which is typically a few gigahertz — a factor of around 1,000 smaller. Femtosecond combs produced by mode-locked lasers typically have a tooth spacing of 100 megahertz to 1 gigahertz. This very tight spacing is also less than ideal, because it reduces the power per comb line, but the spacing can be increased by filtering out a subset of the comb lines. Lowering the spacing of the microresonator combs to this value will be a challenge: it would require increasing the diameter of the toroid to about 0.5 mm, while preserving its extraordinarily high quality factor and simultaneously increasing the laser power to around 5 watts to maintain the nonlinear effect. Overcoming such problems will make for interesting times for this exciting new comb technology. ■

Steven T. Cundiff is at JILA, National Institute of Standards and Technology, University of Colorado, Boulder, Colorado 80309-0440, USA.
e-mail: cundiffs@jila.colorado.edu

1. Jones, D. J. *et al. Science* **288,** 635–639 (2000).
2. Holzwarth, R. *et al. Phys. Rev. Lett.* **85,** 2264–2267 (2000).
3. Cundiff, S. T. & Ye, J. *Rev. Mod. Phys.* **75,** 325–342 (2003).
4. Udem, Th., Holzwarth, R. & Hänsch, T. W. *Nature* **416,** 233–237 (2002).
5. Ma, L. S. *Opt. Photon. News* **18**(9), 42–47 (2007).
6 Thorpe, M. J., Moll, K. D., Jones, R. J., Safdi, B. & Ye, J. *Science* **311,** 1595–1599 (2006).
7. http://nobelprize.org/nobel_prizes/physics/laureates/2005/index.html (2005).
8. Del'Haye, P. *et al. Nature* **450,** 1214–1217 (2007).
9. Kourogi, M., Nakagawa, K. & Ohtsu, M. *IEEE J. Quantum Electron.* **29,** 2693–2701 (1993).
10. Telle, H. R. *et al. App. Phys. B* **69,** 327–332 (1999).
11. Armani, D. K., Kippenberg, T. J., Spillane, S. M. & Vahala, K. J. *Nature* **421,** 925–928 (2003).
12. Ma, L.-S. *et al. Science* **303,** 1843–1845 (2004).

Correction

The photograph accompanying the obituary of Arthur Kornberg by Tania A. Baker (*Nature* **450,** 809; 2007) was inverted left-to-right. Here is the picture in the correct orientation. It shows Dr Kornberg, his first wife Sylvy, and a model of the DNA double-helix (now right-handed, not left-handed).

Stem cell treatment of dystrophic dogs

Arising from: M. Sampaolesi *et al.* *Nature* **444,** 574–579 (2006)

Human muscular dystrophies are devastating and incurable inherited diseases. Hopes of progress towards therapy of muscular dystrophies were aroused when Sampaolesi *et al.*[1] reported "extensive recovery of dystrophin expression, normal muscle…function", and "remarkable clinical amelioration" in golden retriever muscular dystrophy dogs treated with 'mesoangioblasts'. Here I re-examine their results, showing how their assessments might be flawed and their conclusions overstated. Further studies will be required to evaluate fully the clinical potential of this work.

My first concern is that control and test dogs were not matched for disease characteristics at the start of treatment to avoid inadvertently biased groupings in small experimental cohorts, where extensive individual variations exist. Also, the authors' evaluations were not blinded. For example, their Supplementary Videos[1] seem to show greater human encouragement of treated compared with untreated dogs, so functional recovery may not be independently verifiable by viewers, especially the "striking improvement of motility" in the older dogs treated with mesoangioblasts.

Second, Fig. 5a of Sampaolesi *et al.*[1] indicates that muscle strength declined, rather than was maintained as they imply, in the treated legs of all dystrophic dogs in which it was assessed. At later time points, strength was always less than at 5 months. Between the ages of 5 and 9 months, the averaged decline in strength of the two older treated dogs was ~40% (~0.079 to ~0.048 kg^{-1} (%)), even more than the ~25% decline in the control untreated dystrophic dog over the same period. Against this real ~40% decline, purported improvements of ~50% and ~80% in treated leg-muscle strength relative to contralateral, untreated muscles of these two dogs (Fig. 5b of Sampaolesi *et al.*[1]) are misleading and probably explained by simultaneous declines in contralateral muscle strength of ~60% and ~67%, respectively.

Third, the dogs Valgus, Varus and Vaccin each received 5×10^7 mesoangioblasts per treatment, but these were infused into the aortic arch of Valgus and the left femoral artery of the others. At biopsy, ~30–70% of fibres in sections of Valgus' left sartorius and gastrocnemius muscles appeared to be dystrophin-positive (Fig. 4a of Sampaolesi *et al.*[1]), whereas the equivalent dystrophin-positive proportion of Varus' left sartorius was ~0–10%, and that of Vaccin's left gastrocnemius was only 0–5%. Numbers of mesoangioblasts reaching the lower left leg from upper aortic infusion, after major systemic blood diversions[2], should have been ~10 times less than from direct femoral infusion. However, no assessed muscles of Valgus had fewer—and certainly not 10 times fewer—dystrophin-positive fibres than corresponding muscles of Varus or Vaccin. Moreover, some of the untreated muscles of Varus and Vaccin revealed up to 50% dystrophin-positive fibres (Fig. 4a of Sampaolesi *et al.*[1]).

Minimal mesoangioblast recirculation[3] cannot easily explain these anomalies, whereas dystrophin-positive fibres in both treated and untreated muscles might represent false positives or revertants[4], rather than evidence of mesoangioblast engraftment. Control biopsies from pre-treatment and untreated dystrophic dogs[4] would have allowed these possibilities to be differentiated. Alternatively, if recirculation and engraftment is responsible for similar percentages of dystrophin-positive fibres in treated and 'untreated' tibialis cranialis muscles (Fig. 4a of Sampaolesi *et al.*[1]), then contrasting their strengths (Fig. 5b of Sampaolesi *et al.*[1]) is unfounded.

There is an indicator of benefit arising from this trial, although it is, perhaps, due to concurrent immunosuppression rather than to mesoangioblasts. Supplementary Fig. 7 of Sampaolesi *et al.*[1] shows that levels of the muscle-breakdown marker serum creatine kinase decreased markedly soon after initiating immunosuppression (as previously seen in *mdx* mice[5]) and before injection of heterologous mesoangioblasts, whereas in the days immediately after mesoangioblast injection, creatine kinase levels varied randomly, decreasing substantially (>5,000 U) in three instances but increasing in five. This does not support the authors' hypothesis that creatine kinase reductions demonstrate mesoangioblast reconstitution of muscle fibres.

Control dystrophic dogs lived, on average, 129 days longer than six out of ten treated dogs (123 days longer than those treated with autologous mesoangioblasts). Three of the four remaining treated dogs, described as "well" 400 days post natal, "rapidly lost walking ability" when immunosuppression ceased (257 days post natal for two of these).

Altogether, the evidence presented by Sampaolesi *et al.*[1] does not convince me that the dogs benefited from mesoangioblast treatment; a rigorous demonstration correlating muscle function, dystrophin expression and mesoangioblast infusion, with adequate controls, would have been helpful in this regard. It is therefore premature to consider a clinical trial in humans as a justifiable extension of this study.

Allan H. Bretag[1]
[1]Sansom Institute, School of Pharmacy and Medical Sciences, University of South Australia, Adelaide, South Australia 5000, Australia.
e-mail: a.bretag@unisa.edu.au

Received 3 January; accepted 17 October 2007.

1. Sampaolesi, M. *et al.* Mesoangioblast stem cells ameliorate muscle function in dystrophic dogs. *Nature* **444,** 574–579 (2006).
2. Guyton, A. C. *Textbook of Medical Physiology* 7th edn, 230 (Saunders, Philadelphia, 1986).
3. Galvez, B. G. *et al.* Complete repair of dystrophic skeletal muscle by mesoangioblasts with enhanced migration ability. *J. Cell Biol.* **174,** 231–243 (2006).
4. Dell'Agnola, C. *et al.* Hematopoietic stem cell transplantation does not restore dystrophin expression in Duchenne muscular dystrophy dogs. *Blood* **104,** 4311–4318 (2004).
5. De Luca, A. *et al.* A multidisciplinary evaluation of the effectiveness of cyclosporine A in dystrophic mdx mice. *Am. J. Pathol.* **166,** 477–489 (2005).

doi:10.1038/nature06437

Sampaolesi et al. *reply*

Replying to: A. H. Bretag *Nature* **450,** doi:10.1038/nature06437 (2007)

Bretag[1] questions our finding[2] that "dogs benefited from mesoangioblast treatment". We believe that this scepticism is not supported by careful examination of our data[2].

Golden retriever dogs could not be matched for disease characteristics because they were chosen at 1 month of age, before the appearance of clinical symptoms. Indeed, the dog Vaccine, which did not

show any clinical amelioration, had by chance the highest muscle contraction force at the beginning of treatment (Fig. 5a of ref. 2, green triangle). Dogs were encouraged in all cases: the videos of control dogs and of Vaccine lack audio but show the instructor encouraging the dogs with a caress, to which they react by swinging the tail and moving towards the instructor.

Our Fig. 5a (ref. 2) shows that, after a decrease between 5 and 9 months, contraction force increases again only in those dogs treated with donor cells; Fig. 5b of ref. 2, which compares treated and untreated legs of the same animal, shows amelioration of contraction force in the three animals treated with donor cells and not in the one treated with autologous cells. In Duchenne muscular dystrophy, absence of dystrophin causes a reduced force of muscle contraction. We demonstrated a causal relation between dystrophin expression and force recovery, showing (Fig. 5c, d of ref. 2) that in the same muscle, dystrophin-positive fibres have normal force of contraction whereas dystrophin-negative fibres are as weak as untreated, dystrophic fibres[2].

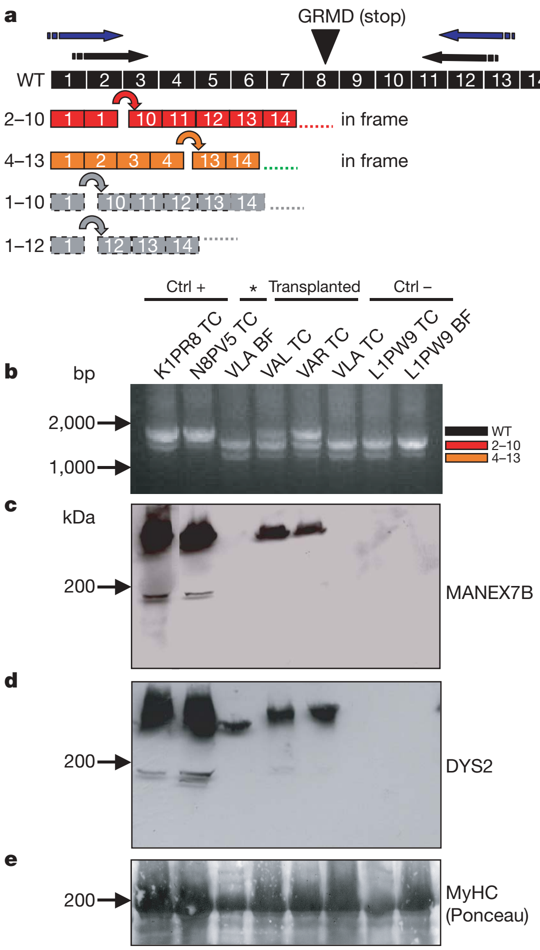


Figure 1 | Expression of wild-type dystrophin in dogs transplanted with wild-type mesoangioblasts. a, Four alternatively processed dystrophin gene transcripts can be amplified in golden retriever muscular dystrophy (GRMD) canine muscle tissue[9]. Only transcripts corresponding to exons 2–10 and 4–13 restore the correct reading frame. Wild type (WT, black) indicates the full-length dystrophin transcript; black and blue arrows indicate the position of oligonucleotides[9] used for the nested polymerase chain reaction (PCR). **b**, Three transcripts are amplified by the nested RT–PCR analysis: the upper band corresponds to wild-type dystrophin (black), and the intermediate and lower bands correspond to the two in-frame transcripts (red and orange, respectively). Note the presence of wild-type transcript in wild-type dogs (K1PR8 and N8PV5) and in dystrophic, transplanted dogs (VAL, Valgus; VAR, Varus), but not in untreated (Ctrl−: VLA and L1PW9) dystrophic dogs. BF, biceps femoralis; TC, tibialis cranialis. **c**, Western blot analysis using MANEX7B antibody (recognizing exons 7/8, a gift from G. Morris) reveals wild-type dystrophin only in transplanted and wild-type dogs. **d**, Western blot analysis using DYS2 antibody (recognizing the carboxy terminus) reveals the wild-type dystrophin as above and a lower molecular mass protein in only one (asterisk; BF, biceps femoralis) muscle of Vlan (VLA), an untreated, dystrophic dog that shows a significant amount of revertant dystrophin without cell transplantation. **e**, Red Ponceau staining of myosin heavy chains of western blot shown in **d**.

Contrary to Bretag's assertion[1], Valgus received about 3×10^7 mesoangioblasts in the left femoralis artery and 2×10^7 in the left subclavian artery. Such a small difference (3×10^7 cells in the case of Valgus versus 5×10^7 cells in standard injections in the femoralis artery) cannot result in a marked difference in the number of dystrophin-positive fibres; the low numbers of dystrophin-positive fibres in Varus' sartorius may be due to more distal insertion of the catheter in the femoral artery, so that injected cells may not have reached the cranial part of the sartorius. Notably, we noticed and pointed out variable distribution of mesoangioblasts in the injected muscles. For this reason, we analysed more than 50 individual biopsies and found extensive, albeit variable, expression of dystrophin in treated dogs.

The presence of dystrophin in the contralateral muscles as well is not due to revertant fibres because an antibody directed against the mutated domain recognizes dystrophin in the muscles of transplanted dogs, confirming that wild-type dystrophin may only derive from wild-type donor cells. Previously unnoticed[3], the biceps femoralis of one untreated dog shows a clear accumulation of revertant dystrophin, easily distinguishable from the wild-type protein (Fig. 1).

It is true that the first drop in serum creatine kinase for Valgus, Varus and Viko followed the first treatment with cyclosporine and is probably caused by the drug, but all the other declines occur after infusion of mesoangioblasts and in the continuous presence of the drug, so cannot be ascribed to it.

The cyclosporine effect is controversial, but Bretag[1] quotes the only claim of benefit from cyclosporine[4] and ignores others claiming deleterious[5–7] effects; moreover, the absence of any beneficial cyclosporine effect in bone-marrow-transplanted, dystrophic dogs[8] is not mentioned.

Long-term survival was not a goal of this study, as immune suppression was suspended at the end of the experiment. Pathologists did not find any feature at autopsy that could have related to mesoangioblast accumulation.

Our results showed extensive dystrophin accumulation, clinical improvement and force preservation in the best available pre-clinical model of Duchenne muscular dystrophy. We are confident that our data are correct and form a suitable basis for future clinical experimentation.

Maurilio Sampaolesi[1], Stephane Blot[2], Roberto Bottinelli[3] & Giulio Cossu[1]
[1]San Raffaele Scientific Institute, Stem Cell Research Institute, Via Olgettina 58, 20132 Milan, Italy.
e-mail: cossu.giulio@hsr.it
[2]Neurobiology Laboratory, Ecòle Veterinaire d'Alfort, 7 Avenue General de Gaulle, 94704 Maisons-Alfort, cedex, France.
[3]Department of Experimental Medicine University of Pavia, Via Forlanini 6-8, 27100 Pavia, Italy.

1. Bretag, A. H. Stem cell treatment of dystrophic dogs. *Nature* **450**, doi:10.1038/nature06437 (2007).
2. Sampaolesi, M. *et al.* Mesoangioblast stem cells ameliorate muscle function in dystrophic dogs. *Nature* **444**, 574–579 (2006).
3. Kornegay, J. *et al.* The cranial sartorius muscle undergoes true hypertrophy in dogs with golden retriever muscular dystrophy. *Neuromuscul. Disord.* **13**, 493–500 (2003).
4. De Luca, A. *et al.* A multidisciplinary evaluation of the effectiveness of cyclosporine A in dystrophic mdx mice. *Am. J. Pathol.* **166**, 477–489 (2005).
5. St-Pierre, S. J. G. *et al.* Glucocorticoid treatment alleviates dystrophic myofiber pathology by activation of the calcineurin/NF-AT pathway. *FASEB J.* **18**, 1937–1939 (2004).

6. Stupka, N. *et al.* The calcineurin signal transduction pathway is essential for successful muscle regeneration in mdx dystrophic mice. *Acta Neuropathol. (Berl.)* **107,** 299–310 (2004).
7. Parsons, S. A. *et al.* Genetic disruption of calcineurin improves skeletal muscle pathology and cardiac disease in a mouse model of limb-girdle muscular dystrophy. *J. Biol. Chem.* **282,** 10068–10078 (2007).
8. Dell'Agnola, C. *et al.* Hematopoietic stem cell transplantation does not restore dystrophin expression in Duchenne muscular dystrophy dogs. *Blood* **104,** 4311–4318 (2004).
9. Schatzberg, S. J. *et al.* Alternative dystrophin gene transcripts in golden retriever muscular dystrophy. *Muscle Nerve* **21,** 991–998 (1998).

doi:10.1038/nature06438

Stem cell treatment of dystrophic dogs

Arising from: M. Sampaolesi *et al.* *Nature* **444,** 574–579 (2006)

Human muscular dystrophies are devastating and incurable inherited diseases. Hopes of progress towards therapy of muscular dystrophies were aroused when Sampaolesi *et al.*[1] reported "extensive recovery of dystrophin expression, normal muscle…function", and "remarkable clinical amelioration" in golden retriever muscular dystrophy dogs treated with 'mesoangioblasts'. Here I re-examine their results, showing how their assessments might be flawed and their conclusions overstated. Further studies will be required to evaluate fully the clinical potential of this work.

My first concern is that control and test dogs were not matched for disease characteristics at the start of treatment to avoid inadvertently biased groupings in small experimental cohorts, where extensive individual variations exist. Also, the authors' evaluations were not blinded. For example, their Supplementary Videos[1] seem to show greater human encouragement of treated compared with untreated dogs, so functional recovery may not be independently verifiable by viewers, especially the "striking improvement of motility" in the older dogs treated with mesoangioblasts.

Second, Fig. 5a of Sampaolesi *et al.*[1] indicates that muscle strength declined, rather than was maintained as they imply, in the treated legs of all dystrophic dogs in which it was assessed. At later time points, strength was always less than at 5 months. Between the ages of 5 and 9 months, the averaged decline in strength of the two older treated dogs was ~40% (~0.079 to ~0.048 kg^{-1} (%)), even more than the ~25% decline in the control untreated dystrophic dog over the same period. Against this real ~40% decline, purported improvements of ~50% and ~80% in treated leg-muscle strength relative to contralateral, untreated muscles of these two dogs (Fig. 5b of Sampaolesi *et al.*[1]) are misleading and probably explained by simultaneous declines in contralateral muscle strength of ~60% and ~67%, respectively.

Third, the dogs Valgus, Varus and Vaccin each received 5×10^7 mesoangioblasts per treatment, but these were infused into the aortic arch of Valgus and the left femoral artery of the others. At biopsy, ~30–70% of fibres in sections of Valgus' left sartorius and gastrocnemius muscles appeared to be dystrophin-positive (Fig. 4a of Sampaolesi *et al.*[1]), whereas the equivalent dystrophin-positive proportion of Varus' left sartorius was ~0–10%, and that of Vaccin's left gastrocnemius was only 0–5%. Numbers of mesoangioblasts reaching the lower left leg from upper aortic infusion, after major systemic blood diversions[2], should have been ~10 times less than from direct femoral infusion. However, no assessed muscles of Valgus had fewer—and certainly not 10 times fewer—dystrophin-positive fibres than corresponding muscles of Varus or Vaccin. Moreover, some of the untreated muscles of Varus and Vaccin revealed up to 50% dystrophin-positive fibres (Fig. 4a of Sampaolesi *et al.*[1]).

Minimal mesoangioblast recirculation[3] cannot easily explain these anomalies, whereas dystrophin-positive fibres in both treated and untreated muscles might represent false positives or revertants[4], rather than evidence of mesoangioblast engraftment. Control biopsies from pre-treatment and untreated dystrophic dogs[4] would have allowed these possibilities to be differentiated. Alternatively, if recirculation and engraftment is responsible for similar percentages of dystrophin-positive fibres in treated and 'untreated' tibialis cranialis muscles (Fig. 4a of Sampaolesi *et al.*[1]), then contrasting their strengths (Fig. 5b of Sampaolesi *et al.*[1]) is unfounded.

There is an indicator of benefit arising from this trial, although it is, perhaps, due to concurrent immunosuppression rather than to mesoangioblasts. Supplementary Fig. 7 of Sampaolesi *et al.*[1] shows that levels of the muscle-breakdown marker serum creatine kinase decreased markedly soon after initiating immunosuppression (as previously seen in *mdx* mice[5]) and before injection of heterologous mesoangioblasts, whereas in the days immediately after mesoangioblast injection, creatine kinase levels varied randomly, decreasing substantially (>5,000 U) in three instances but increasing in five. This does not support the authors' hypothesis that creatine kinase reductions demonstrate mesoangioblast reconstitution of muscle fibres.

Control dystrophic dogs lived, on average, 129 days longer than six out of ten treated dogs (123 days longer than those treated with autologous mesoangioblasts). Three of the four remaining treated dogs, described as "well" 400 days post natal, "rapidly lost walking ability" when immunosuppression ceased (257 days post natal for two of these).

Altogether, the evidence presented by Sampaolesi *et al.*[1] does not convince me that the dogs benefited from mesoangioblast treatment; a rigorous demonstration correlating muscle function, dystrophin expression and mesoangioblast infusion, with adequate controls, would have been helpful in this regard. It is therefore premature to consider a clinical trial in humans as a justifiable extension of this study.

Allan H. Bretag[1]
[1]Sansom Institute, School of Pharmacy and Medical Sciences, University of South Australia, Adelaide, South Australia 5000, Australia.
e-mail: a.bretag@unisa.edu.au

Received 3 January; accepted 17 October 2007.

1. Sampaolesi, M. *et al.* Mesoangioblast stem cells ameliorate muscle function in dystrophic dogs. *Nature* **444,** 574–579 (2006).
2. Guyton, A. C. *Textbook of Medical Physiology* 7th edn, 230 (Saunders, Philadelphia, 1986).
3. Galvez, B. G. *et al.* Complete repair of dystrophic skeletal muscle by mesoangioblasts with enhanced migration ability. *J. Cell Biol.* **174,** 231–243 (2006).
4. Dell'Agnola, C. *et al.* Hematopoietic stem cell transplantation does not restore dystrophin expression in Duchenne muscular dystrophy dogs. *Blood* **104,** 4311–4318 (2004).
5. De Luca, A. *et al.* A multidisciplinary evaluation of the effectiveness of cyclosporine A in dystrophic mdx mice. *Am. J. Pathol.* **166,** 477–489 (2005).

doi:10.1038/nature06437

Sampaolesi et al. *reply*

Replying to: A. H. Bretag *Nature* **450,** doi:10.1038/nature06437 (2007)

Bretag[1] questions our finding[2] that "dogs benefited from mesoangioblast treatment". We believe that this scepticism is not supported by careful examination of our data[2].

Golden retriever dogs could not be matched for disease characteristics because they were chosen at 1 month of age, before the appearance of clinical symptoms. Indeed, the dog Vaccine, which did not

show any clinical amelioration, had by chance the highest muscle contraction force at the beginning of treatment (Fig. 5a of ref. 2, green triangle). Dogs were encouraged in all cases: the videos of control dogs and of Vaccine lack audio but show the instructor encouraging the dogs with a caress, to which they react by swinging the tail and moving towards the instructor.

Our Fig. 5a (ref. 2) shows that, after a decrease between 5 and 9 months, contraction force increases again only in those dogs treated with donor cells; Fig. 5b of ref. 2, which compares treated and untreated legs of the same animal, shows amelioration of contraction force in the three animals treated with donor cells and not in the one treated with autologous cells. In Duchenne muscular dystrophy, absence of dystrophin causes a reduced force of muscle contraction. We demonstrated a causal relation between dystrophin expression and force recovery, showing (Fig. 5c, d of ref. 2) that in the same muscle, dystrophin-positive fibres have normal force of contraction whereas dystrophin-negative fibres are as weak as untreated, dystrophic fibres[2].

Contrary to Bretag's assertion[1], Valgus received about 3×10^7 mesoangioblasts in the left femoralis artery and 2×10^7 in the left subclavian artery. Such a small difference (3×10^7 cells in the case of Valgus versus 5×10^7 cells in standard injections in the femoralis artery) cannot result in a marked difference in the number of dystrophin-positive fibres; the low numbers of dystrophin-positive fibres in Varus' sartorius may be due to more distal insertion of the catheter in the femoral artery, so that injected cells may not have reached the cranial part of the sartorius. Notably, we noticed and pointed out variable distribution of mesoangioblasts in the injected muscles. For this reason, we analysed more than 50 individual biopsies and found extensive, albeit variable, expression of dystrophin in treated dogs.

The presence of dystrophin in the contralateral muscles as well is not due to revertant fibres because an antibody directed against the mutated domain recognizes dystrophin in the muscles of transplanted dogs, confirming that wild-type dystrophin may only derive from wild-type donor cells. Previously unnoticed[3], the biceps femoralis of one untreated dog shows a clear accumulation of revertant dystrophin, easily distinguishable from the wild-type protein (Fig. 1).

It is true that the first drop in serum creatine kinase for Valgus, Varus and Viko followed the first treatment with cyclosporine and is probably caused by the drug, but all the other declines occur after infusion of mesoangioblasts and in the continuous presence of the drug, so cannot be ascribed to it.

The cyclosporine effect is controversial, but Bretag[1] quotes the only claim of benefit from cyclosporine[4] and ignores others claiming deleterious[5–7] effects; moreover, the absence of any beneficial cyclosporine effect in bone-marrow-transplanted, dystrophic dogs[8] is not mentioned.

Long-term survival was not a goal of this study, as immune suppression was suspended at the end of the experiment. Pathologists did not find any feature at autopsy that could have related to mesoangioblast accumulation.

Our results showed extensive dystrophin accumulation, clinical improvement and force preservation in the best available pre-clinical model of Duchenne muscular dystrophy. We are confident that our data are correct and form a suitable basis for future clinical experimentation.

Maurilio Sampaolesi[1], Stephane Blot[2], Roberto Bottinelli[3] & Giulio Cossu[1]
[1]San Raffaele Scientific Institute, Stem Cell Research Institute, Via Olgettina 58, 20132 Milan, Italy.
e-mail: cossu.giulio@hsr.it
[2]Neurobiology Laboratory, Ecòle Veterinaire d'Alfort, 7 Avenue General de Gaulle, 94704 Maisons-Alfort, cedex, France.
[3]Department of Experimental Medicine University of Pavia, Via Forlanini 6-8, 27100 Pavia, Italy.

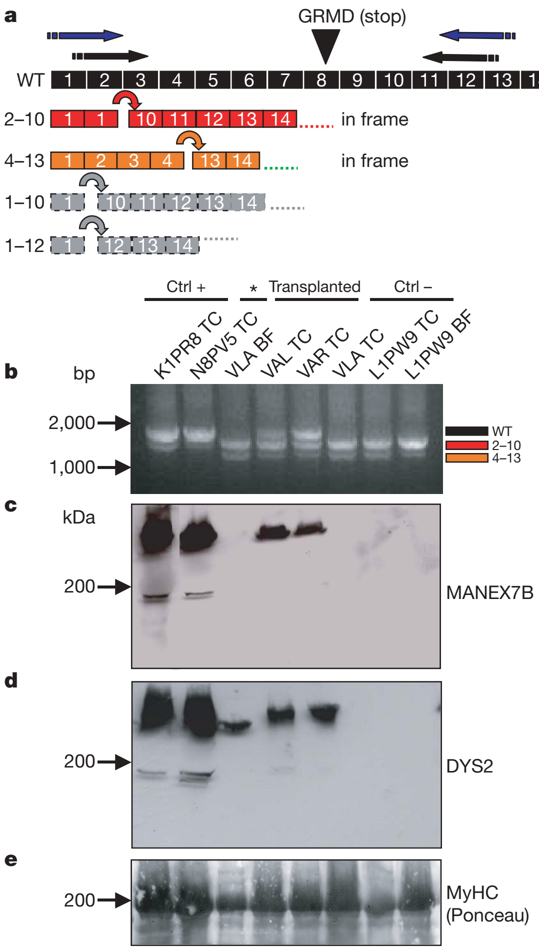


Figure 1 | Expression of wild-type dystrophin in dogs transplanted with wild-type mesoangioblasts. **a**, Four alternatively processed dystrophin gene transcripts can be amplified in golden retriever muscular dystrophy (GRMD) canine muscle tissue[9]. Only transcripts corresponding to exons 2–10 and 4–13 restore the correct reading frame. Wild type (WT, black) indicates the full-length dystrophin transcript; black and blue arrows indicate the position of oligonucleotides[9] used for the nested polymerase chain reaction (PCR). **b**, Three transcripts are amplified by the nested RT–PCR analysis: the upper band corresponds to wild-type dystrophin (black), and the intermediate and lower bands correspond to the two in-frame transcripts (red and orange, respectively). Note the presence of wild-type transcript in wild-type dogs (K1PR8 and N8PV5) and in dystrophic, transplanted dogs (VAL, Valgus; VAR, Varus), but not in untreated (Ctrl−: VLA and L1PW9) dystrophic dogs. BF, biceps femoralis; TC, tibialis cranialis. **c**, Western blot analysis using MANEX7B antibody (recognizing exons 7/8, a gift from G. Morris) reveals wild-type dystrophin only in transplanted and wild-type dogs. **d**, Western blot analysis using DYS2 antibody (recognizing the carboxy terminus) reveals the wild-type dystrophin as above and a lower molecular mass protein in only one (asterisk; BF, biceps femoralis) muscle of Vlan (VLA), an untreated, dystrophic dog that shows a significant amount of revertant dystrophin without cell transplantation. **e**, Red Ponceau staining of myosin heavy chains of western blot shown in **d**.

1. Bretag, A. H. Stem cell treatment of dystrophic dogs. *Nature* **450**, doi:10.1038/nature06437 (2007).
2. Sampaolesi, M. *et al.* Mesoangioblast stem cells ameliorate muscle function in dystrophic dogs. *Nature* **444**, 574–579 (2006).
3. Kornegay, J. *et al.* The cranial sartorius muscle undergoes true hypertrophy in dogs with golden retriever muscular dystrophy. *Neuromuscul. Disord.* **13**, 493–500 (2003).
4. De Luca, A. *et al.* A multidisciplinary evaluation of the effectiveness of cyclosporine A in dystrophic mdx mice. *Am. J. Pathol.* **166**, 477–489 (2005).
5. St-Pierre, S. J. G. *et al.* Glucocorticoid treatment alleviates dystrophic myofiber pathology by activation of the calcineurin/NF-AT pathway. *FASEB J.* **18**, 1937–1939 (2004).

6. Stupka, N. *et al.* The calcineurin signal transduction pathway is essential for successful muscle regeneration in mdx dystrophic mice. *Acta Neuropathol. (Berl.)* **107,** 299–310 (2004).
7. Parsons, S. A. *et al.* Genetic disruption of calcineurin improves skeletal muscle pathology and cardiac disease in a mouse model of limb-girdle muscular dystrophy. *J. Biol. Chem.* **282,** 10068–10078 (2007).
8. Dell'Agnola, C. *et al.* Hematopoietic stem cell transplantation does not restore dystrophin expression in Duchenne muscular dystrophy dogs. *Blood* **104,** 4311–4318 (2004).
9. Schatzberg, S. J. *et al.* Alternative dystrophin gene transcripts in golden retriever muscular dystrophy. *Muscle Nerve* **21,** 991–998 (1998).

doi:10.1038/nature06438

REVIEWS

Superconductivity without phonons

P. Monthoux[1,2], D. Pines[3,4] & G. G. Lonzarich[5]

The idea of superconductivity without the mediating role of lattice vibrations (phonons) has a long history. It was realized soon after the publication of the Bardeen–Cooper–Schrieffer (BCS) theory of superconductivity 50 years ago that a full treatment of both the charge and spin degrees of freedom of the electron predicts the existence of attractive components of the effective interaction between electrons even in the absence of lattice vibrations—a particular example is the effective interaction that depends on the relative spins of the electrons. Such attraction without phonons can lead to electronic pairing and to unconventional forms of superconductivity that can be much more sensitive than traditional (BCS) superconductivity to the precise details of the crystal structure and to the electronic and magnetic properties of a material.

Superconductivity, the dissipationless flow of electrical current, is a striking manifestation of a subtle form of quantum rigidity on the macroscopic scale. In such a quantum state, the low-energy behaviour is insensitive to many perturbations, and is typically characterized by a small number of parameters that can be determined experimentally but are difficult to calculate accurately from first principles[1]. The major achievement of Bardeen, Cooper and Schrieffer was to identify the fundamental organizing principle responsible for superconductivity, and show in a simple model calculation how the low-lying excited states could be characterized by three macroscopic parameters that were easily determined from experiment[2,3].

In real metals at temperatures that are low, but above the superconducting state, the experimental observations can be accounted for in terms of a low density gas of thermally excited particles (and antiparticles) with the same charge and spin, but not the same mass, as the electron (and positron). These particles are referred to as quasiparticles. In the hypothetical absence of the electron–electron and electron–ion interactions, these quasiparticles reduce to bare electrons. In real metals, one may think of a quasiparticle as an electron plus a co-moving screening cloud. In extreme cases, such as the heavy electron superconductors that are of particular interest in this Review, the quasiparticles can behave as if they have, for example, effective masses two or more orders of magnitude greater than that of electrons and hence move relatively slowly[4–8]. The low velocity of the quasiparticles reflects the fact that the transport of charge in these systems is strongly impeded by the lattice potential and collisions with other particles. We note that the quasiparticle picture described here appears to have a wide range of applicability—remarkably, even on the border of, but not too close to, instabilities.

At a deep level, the electron–electron interaction is an induced interaction: that is, one electron (the polarizer) induces a change in the electromagnetic field in free space and the second electron (the analyser) samples this change. One can also think of the interaction between quasiparticles in the same manner, but the electromagnetic field now also includes that produced by the other particles in the material. We refer to the total electromagnetic field seen by the quasiparticles as the effective field. This suggests that the quasiparticle interaction can be very different from that of electrons, and is the key to understanding superconductivity. In particular, the charge–charge interaction between quasiparticles can have attractive regions in space owing to the collective effects of the electron and lattice system. Importantly, the quasiparticles also experience an effective spin–spin interaction (to be discussed below) that can be several orders of magnitude stronger than the conventional magnetic dipole interaction between bare electrons.

Where the quasiparticle interaction has attractive regions, a new kind of quantum order can emerge in which the quasiparticles form pairs locked into a coherent state extending over macroscopic dimensions. This is the phenomenon of superconductivity.

In conventional superconductors (for example, Al, Pb and Hg), the effective charge–charge interaction between quasiparticles is dominant. The attractive part of the interaction in these cases arises from the induced effects of quantized lattice vibrations or phonons. In this Review, we begin with a primer on electron–phonon-mediated superconductivity and then explain how quasiparticle pairs can form via the above mentioned spin–spin interaction that does not involve phonons explicitly and depends on the spins of the interacting quasiparticles rather than their charges.

Phonon-mediated superconductivity

For the first 40 years after the discovery of superconductivity in 1911 (ref. 9), those seeking a microscopic mechanism for superconductivity did not consider the possibility that phonons could play any part. Rather, the focus was on a possible special role for the crystal structure of the lattice[10] or on the possibility that, if one treated properly the consequences of the Coulomb interaction between the electrons[11,12], superconductivity might emerge. Matters changed when researchers working at Rutgers University and the National Bureau of Standards discovered the isotope effect—that in lead, the transition temperature below which superconductivity was found depended inversely on the square root of the isotopic mass[13,14]. This discovery motivated Frohlich[15] and Bardeen[16] to explore the possibility that it was the change in the quasiparticle energy spectrum due to interactions with the phonons that led to superconductivity, an approach that soon turned out be to fruitless.

Frohlich had also suggested[17] that somehow the attractive part of the phonon-induced interaction between quasiparticles might play a part. However, this proposal also seemed flawed, as it appeared that the much stronger repulsive direct Coulomb interaction would swamp the effect. (Evgeny Lif'shitz mentioned to one of us that Landau said at one point "You can't repeal Coulomb's law".) This objection turned out not to be insuperable, as a few years later, Bardeen and one of us[18] were able to show, in a simple model

[1]School of Physics, [2]Centre for Science at Extreme Conditions, University of Edinburgh, Edinburgh EH9 3JZ, UK. [3]G756 Los Alamos National Laboratory, Los Alamos, New Mexico 87544, USA. [4]Institute for Complex Adaptive Matter and Physics Department, University of California at Davis, One Shields Avenue, Davis, California 95616, USA. [5]Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK.

calculation, that when full account was taken of screening[19] by the electron and lattice system as a whole one could still get a regime in space and time in which the interaction between quasiparticles is attractive.

Cooper[20] then showed how this effective interaction between quasiparticles could give rise to the formation of quasiparticle pairs (see Fig. 1a for a real space illustration of an *s*-wave Cooper-pair state). Within a year, Bardeen, Cooper and Schrieffer developed a simple microscopic description of the superconducting ground state in terms of Cooper pairs.

An effective attractive component of the overall interaction between quasiparticles was thereby established as the key to a superconducting transition. For the next nearly 30 years, the primacy of the phonon-induced attraction—with a predicted ceiling to the superconducting transition temperature (T_c) below the boiling point of liquid nitrogen—appeared to be well established. Matters changed with the discovery by Bednorz and Müller[21] of the first member of the copper oxide family of superconductors, which have T_c values of up to 160 K (twice the boiling point of liquid nitrogen). These unexpectedly high T_c values and other anomalous properties of the copper oxides led theorists to explore in detail the possibility of superconductivity without phonons, that is, of an effective attractive interaction between quasiparticles that does not originate in their coupling to lattice vibrations.

This search had begun earlier, in particular with the suggestion that a spin–spin quasiparticle interaction[22], analogous to that thought to stabilize superfluidity in liquid ^{3}He (refs 23, 24), might also be relevant to superconductivity on the border of itinerant-electron magnetism[25], where the spin–spin interaction described in more detail in the next section is expected to be important. Superconductivity on the border of ferromagnetism[25] has proved to be an elusive phenomenon, and has been observed only relatively recently following a long search[26,27]. However, for nearly a decade before the Bednorz and Mueller discovery, evidence had been accumulating for superconductivity on the border of antiferromagnetism in low-dimensional organic metals[28] and a family of 'heavy electron' compounds[4–7,29–32].

Whereas these materials on the boundary of antiferromagnetism superconduct only at comparatively low temperatures, both their normal state and superconducting properties seemed quite different from those of the conventional metallic superconductors. This led to speculations that the spin–spin quasiparticle interaction, extended from the case of a nearly ferromagnetic metal to the seemingly more subtle case of a nearly antiferromagnetic system, might be the dominant mechanism for quasiparticle pairing in these materials[33–35]. It is this possibility for superconductivity without phonons, both in the nearly ferromagnetic and the nearly antiferromagnetic metals, that we consider in this Review.

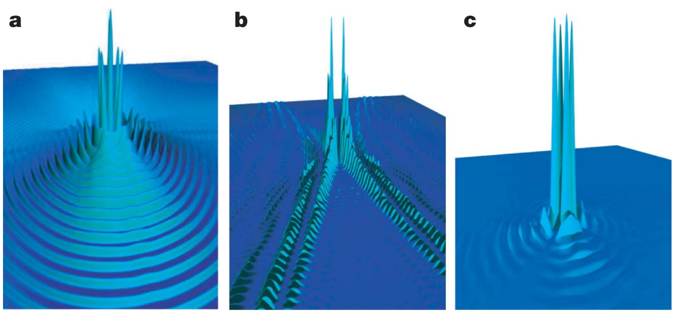


Figure 1 | Cooper-pair states in real space in two dimensions. a–c, The probability of finding one quasiparticle in a Cooper-pair state given that the other partner is at the origin. The symmetric form in **a** is characteristic of an *s*-wave spin-singlet state. The two-fold symmetry in **b** is characteristic of one of the possible *p*-wave spin-triplet states, whereas the four-fold symmetry in **c** is characteristic of a *d*-wave spin-singlet state. The Cooper-pair states shown are predicted for the model interaction described in the text just below the superconducting transition temperature, T_c. In **a** we have used a simple model of the phonon-mediated interaction[3] and in **b** and **c** realistic models of the electronic structure of the ruthenates (**b**) and the copper oxides (**c**)[51]. In the *s*-wave and *p*-wave cases (**a** and **b**, respectively), the probability distribution function extends over many lattice sites, which is the signature of a weakly bound large-sized Cooper pair reflecting a low T_c. In contrast, in the *d*-wave case (**c**), the probability distribution is strongly peaked at the nearest neighbour lattice site to the origin and quickly decays with distance. This indicates a small Cooper-pair size and correspondingly high T_c.

Density and magnetically mediated superconductivity

In this section, we describe a simple framework for thinking about quasiparticle interactions of which the phonon-mediated interaction is a special case. We consider the case in which a quasiparticle at some point $\mathbf{r}$ and time t would see two effective fields, one proportional to the charge density $n(\mathbf{r},t)$ coupled to the quasiparticle charge e, and the other proportional to the spin density or magnetization $\mathbf{m}(\mathbf{r},t)$ coupled to the quasiparticle spin $\mathbf{s}$. In the case of a bare electron, these fields would reduce to the electric and magnetic fields, respectively, that obey Maxwell's equations. In principle, there will also be analogous fields coupled to the quasiparticle current and other quasiparticle properties that will not be considered further here[36,37]. We implicitly assume that the mean value of the density has been subtracted out, and that the system is not magnetically ordered.

In this phenomenological model, the interaction of the given quasiparticle with the fields can be expressed in the form

$$-e[g_n\, n(\mathbf{r},t)] - \mathbf{s}\bullet[g_m\, \mathbf{m}(\mathbf{r},t)] \qquad (1)$$

where g_n and g_m are empirically determined parameters that measure the strength of the coupling to the charge and to the spin, respectively. Note that in electromagnetism the relevant coupling parameters are also obtained experimentally. More generally, the fields seen by the quasiparticles may depend not only on the local densities, but also on the densities in the neighbourhood and at slightly earlier times. Mathematically, the above relation is then replaced by a convolution in space and time. An example of this is the polarization potential approach introduced by one of us for the helium liquids[38].

In quantum electrodynamics, the effective electron–electron interaction including quantum corrections can be calculated by averaging over the statistical distribution of the Maxwell field. In an analogous way, one can obtain the low-frequency quasiparticle–quasiparticle interaction by averaging over the appropriate joint statistical distribution of effective fields. This distribution is likely to be very complex, and thus the usefulness of the effective field picture rests on the fact that it works—namely, that simple approximations can yield significant physical insights into the behaviour of correlated quantum matter.

An intuitive picture of the effective interaction starting with the spin–spin interaction can be given as follows. The spin $\mathbf{s}'$, the polarizer, generates a bare interaction field, taken for simplicity to be localized at the origin in space and at the origin in time, of strength $g_m\mathbf{s}'$. In the linear response approximation, this field induces a magnetization given by $\mathbf{m}(\mathbf{r},t) = g_m\mathbf{s}'\chi_m(\mathbf{r},t)$, where χ_m is the non-local magnetic susceptibility. This induced magnetization produces an induced effective field $g_m\mathbf{m}(\mathbf{r},t)$ that acts on spin $\mathbf{s}'$, the analyser. The induced interaction takes the familiar form of a magnetic moment in an external field given by the second term in equation (1), which in this case reduces to $-g_m^2\mathbf{s}\bullet\mathbf{s}'\chi_m(\mathbf{r},t)$. A similar argument applied to the charge–charge interaction yields an induced interaction of the form $-ee'g_n^2\chi_n(\mathbf{r},t)$, where χ_n is the non-local charge density susceptibility and $e = e'$ is the charge of the interacting quasiparticles. This leads to an overall induced interaction:

$$V_{\mathrm{ind}}(\mathbf{r},t) = -\,e\,e'\,g_n^{\,2}\,\chi_n(\mathbf{r},\,t) - \mathbf{s}\bullet\mathbf{s}'\,g_m^{\,2}\,\chi_m(\mathbf{r},t) \qquad (2)$$

In our model, this induced interaction is in addition to a residual short-range repulsion that prevents charge carriers from coming to

the same place at the same time. An extension of this treatment to the 'strong coupling' limit, where the effective fields are not weak, is discussed below.

For comparison between the predictions of the model interaction (equation (2)) and experiment, it is best to use as inputs the experimentally determined susceptibilities χ_n and χ_m and molecular field constants g_n and g_m. To explore in a qualitative way the range of possible behaviours, it is fruitful, however, to consider simple theoretical models for the susceptibilities in equation (2).

To a first approximation, χ_n and χ_m could be replaced by the susceptibility of non-interacting electrons in a periodic crystal potential, namely, the Lindhard function, which can be obtained by means of standard energy band theory. Although this approximation leads one to understand how the effective interaction between quasiparticles can have attractive regions in space and time, even when phonons are not included, it normally leads to values of T_c so low as to make the experimental observation practically impossible[11,12,39].

The values of T_c could potentially rise to the experimentally accessible range if the density or magnetic susceptibilities χ_n and χ_m (and hence the attractive parts in equation (2)) were enhanced over that of the Lindhard function. Such considerations would naturally lead one to look for superconductivity on the border of density or magnetic instabilities. An improved representation of χ_n and χ_m could be obtained in the random phase approximation[19], a generalization of the Weiss-Stoner model for the magnetic susceptibility, which yields χ_n and χ_m in terms of the Lindhard functions and the effective field constants g_n and g_m. A still better approximation is obtained when these starting effective field constants are replaced by effective temperature dependent ones determined by a self-consistency requirement[40,41].

In contrast to the random phase approximation, this latter self-consistent renormalization model yields values of ferromagnetic or antiferromagnetic transitions that are generally in good agreement with experiment[42]. In particular, this model is in keeping with the suppression of long-range magnetic order at finite temperatures expected for systems in one or two dimensions.

Apart from the molecular field constants, the model interaction in the self-consistent renormalization approximation is then completely specified by the underlying energy band structure that defines the Lindhard functions.

Where to look

The problem of quasiparticle pairing via equation (2) is made subtle by the fact that the interaction will in general have both attractive and repulsive regions in space and time. The issue is then whether a Cooper-pair wavefunction can be constructed that has a large amplitude in the space-time regions where the interaction is attractive and a small amplitude elsewhere. An additional complication compared to the more conventional particle pairing in free space, where all momentum states are available for optimizing the pair waveforms, is the following. In free space, the low-energy electron and positron excitations have momentum close to zero. In the case of a finite density of electrons, the corresponding low energy particle and antiparticle states have a finite momentum. The momenta of zero energy excitations determine a surface in momentum space referred to as the Fermi surface. In the present case, the available states for the pair states are generally restricted to the neighbourhood of the Fermi surface. This implies that the pair wavefunction oscillates in space with wavevector connected with the diameter of the Fermi surface. Because the interaction is dynamical, the appropriate Cooper state will also exhibit a non-trivial time dependence.

In order for such an interaction to bring about quasiparticle pairing, the allowed oscillatory nature of the Cooper state, restricted by the Fermi surface, must match that of the interaction.

In simple metals, the effective charge–charge interaction is dominant; it is repulsive at the origin ($\mathbf{r} = t = 0$) and becomes attractive only at finite times ($\mathbf{r} = 0$, $t > 0$) owing to the effects of phonons (Fig. 2a, b). This then leads to a Cooper state with maximum amplitude at finite time and with zero relative angular momentum (*s*-wave, Fig. 1a). The Pauli principle requires that the spins of the paired quasiparticles be in the singlet state. This time-delayed charge–charge interaction is the key to understanding the conventional superconductors, but may only be part of the story in strongly correlated electron systems.

Particularly in systems close to magnetic instabilities (for example, close to ferromagnetism or antiferromagnetism), the spin susceptibility $\chi_m(\mathbf{r},t)$ can be appreciable, and thus we expect the spin–spin interaction (the second term in equation (2) to be important. A qualitatively new feature of this interaction is that contrary to the charge–charge interaction where $ee' = e^2$ is always positive, the inner products of the two spins can be positive or negative depending on the relative spin orientations of the interacting quasiparticles.

On the border of a ferromagnetic instability, $\chi_m(\mathbf{r},t)$ is maximum and positive at the origin and decays slowly in space and time. Therefore, the spin interaction is attractive for parallel spins, that is, for the spin-triplet state (Fig. 1c). The Pauli principle then requires the Cooper state to have odd angular momentum. Generally, the attraction is strongest in the lowest allowed angular momentum state, or the *p*-wave state (Fig. 1b).

On the other hand, on the border of antiferromagnetism, $\chi_m(\mathbf{r},t)$ is also maximum and positive at the origin and tends to decay in time, but more crucially it oscillates in space with a period comparable to the lattice spacing (Figs 2d, 3). For either the spin-singlet or spin-triplet states, this interaction will have both repulsive and attractive regions in space that tend to produce cancelling effects. It is still possible to get a superconducting instability, provided that the spatial

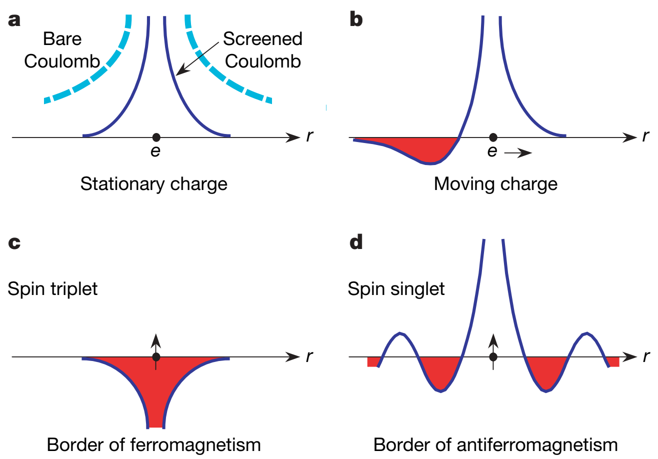


Figure 2 | Schematic illustration of the charge-charge and spin-spin quasiparticle interaction in a metal. Panels **a** and **b** depict the charge–charge interaction potential versus distance (r) produced by a test charge at rest in **a** and moving to the right in **b**. The dashed curve in **a** represents the static Coulomb repulsion and the solid line represents this interaction as screened by the effects of other charged particles (ions and electrons) in the material. We note that in this simple model the interaction is everywhere repulsive in the static limit. When the test charge is moving, however, the interaction develops an attractive region directly behind it (**b**). This attraction is caused by the polarization of the positively charged and heavy ions that lie behind the moving charge and its electron screening cloud. A similar phenomenon arises in the classical description of an object moving through a medium. For example, a fast moving car creates a low pressure (attractive) region just behind it. Panels **c** and **d** represent effective spin–spin interaction potentials set up by a test spin at rest in a material on the border of long-range magnetic order. The interaction potential shown is attractive for two quasiparticles with the same spin (spin triplet state) on the border of ferromagnetism (**c**) and is oscillatory in space on the border of antiferromagnetism (**d**). The illustration in **d** is for the case where the two interacting quasiparticles have opposite spins (spin singlet state). In real materials, the interaction potential is expected to display a complex oscillatory pattern both on the border of ferromagnetism and antiferromagnetism.

variation of the Cooper state can be properly adjusted to match the oscillations of the spin interaction. Generally, the instability is found in the spin-singlet state, which must necessarily have even angular momentum. As the interaction is repulsive at the origin, a non-zero angular momentum state, typically a *d*-wave state, is favoured (Fig. 1c).

We note that the analogues of the ferromagnetic and antiferromagnetic instabilities can exist in the density (described by $\chi_n(\mathbf{r},\mathbf{t})$) as well[43]. The pattern of oscillations in $\chi_n(\mathbf{r},t)$ is determined by the wavevector at which the density response is maximum, and this too can lead to unconventional pairing states. In contrast to the magnetic interaction where the sign of the interaction is different for spin-triplet and spin-singlet Cooper pairs, the density interaction is oblivious to the spin state of the Cooper pairs. Hence, the magnetic interaction offers more freedom to match the attractive regions of the oscillations of the interactions with the quasiparticle states near the Fermi surface available to construct a Cooper-pair wavefunction.

Some surprises

On the border of ferromagnetism the induced interaction is purely attractive at short distances, whereas on the border of antiferromagnetism it is likely to have repulsive regions when the two interacting quasiparticles are close to each other. It might therefore be expected that the search for exotic pairing states on the former border would have proved the more fruitful. This seemed to have been borne out by the discovery of the superfluidity of liquid ^{3}He in which the uniform magnetic susceptibility is strongly enhanced[44]. Given the abundance of metals that exhibit strong ferromagnetic correlations, it is more than a little surprising that it took a quarter of a century to discover a superconducting analogue of liquid ^{3}He, namely, the layered perovskite Sr_2RuO_4 that has a T_c value two orders of magnitude below that of the high T_c copper oxides[26].

Even more perplexing on the other hand is the fact that many examples of superconductivity on the border of antiferromagnetism have been found in the intervening period[4–7,28–32]. These findings would suggest at first sight that the magnetic interaction is giving us little or no insight on where to look for exotic forms of superconductivity.

However, the great lesson of the past decade is that subtleties in the magnetic interaction model only come to the surface after a careful examination of its properties. Although the idea of a magnetic interaction goes back nearly half a century, the computer algorithms and hardware necessary for an exploration of the detailed predictions of the model have only become available more recently.

The results of these theoretical investigations have led to an intuitive understanding of the following: (1) why superconductivity can be particularly robust on the border of antiferromagnetism in a quasi-two-dimensional tetragonal system with high characteristic spin fluctuation frequencies[30,45–48]; (2) how the charge–charge and spin–spin interactions can in some cases work coherently to stabilize anisotropic Cooper-pair states; and (3) why pairing on the border of ferromagnetism is hampered by quite a number of effects, and may depend on subtle details of the electronic structure, that is, features (absent in liquid ^{3}He) of the energy band of the periodic crystal potential[49–51]. Illustrations of these ideas are given below.

The first reason for the robustness of pairing in the presence of antiferromagnetic correlations in a tetragonal structure is that the amplitude of the oscillations in the interaction is strong because of the low dimensionality. The energy density of the interaction waves created by the polarizer falls off more gradually in two dimensions (as 1/distance) than in three dimensions (as 1/distance^2). The second reason is that the repulsive regions of the interaction in real space are along the diagonals of the lattice given that one quasiparticle is at the origin (see Fig. 3). In this case, the crystal symmetry allows one to choose a *d*-wave Cooper state with nodes along the diagonals, thereby neutralizing most of the repulsive regions while retaining the attractive regions. One can easily imagine that it will not always be possible to choose a Cooper-pair state in such an optimal way, and that the initial impression that the oscillations of the interaction are detrimental to superconductivity may only be wrong in special cases. In particular, as the tetragonal structure becomes more and more isotropic under otherwise similar conditions, the model predicts a decrease in the robustness of the pairing. The range in temperature and pressure over which superconductivity is observed was increased by about one order of magnitude in going from cubic $CeIn_3$ (refs 30, 52) to its tetragonal analogues $CeMIn_5$, where M stands for Rh, Ir or Co (refs 31, 32, 53–56; Fig. 4), as anticipated by the magnetic interaction model.

Another case where subtle features of the model considered here could explain puzzling superconducting properties is the first of the heavy-fermion superconductors, $CeCu_2Si_2$ (ref. 4), and the related

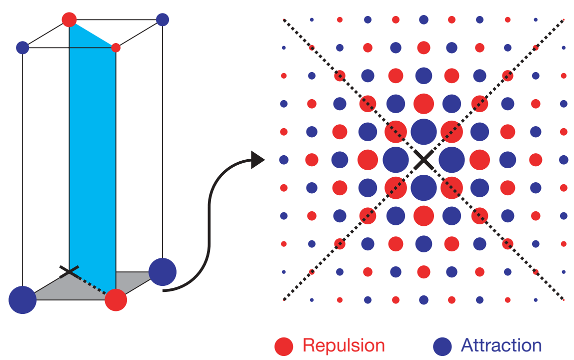


Figure 3 | Magnetic interaction potential in a lattice. Graphical representation of the static magnetic interaction potential in real space seen by a quasiparticle moving on a square crystal lattice given that the other quasiparticle is at the origin (denoted by a cross). The spins of the interacting quasiparticles are taken to be antiparallel, such that the total spin of the Cooper pair is zero. The dashed lines show the regions where the *d*-wave Cooper-pair state has vanishing amplitude. This is the state that best matches the oscillations of the potential, in that a quasiparticle has minimal probability of being on lattice sites when the potential induced by the quasiparticle at the origin is repulsive. The size of the circle in each lattice site is a representation of the absolute magnitude of the potential (on a logarithmic scale). This picture is appropriate for a system on the border of antiferromagnetism in which the period of the real space oscillations of the potential is precisely commensurate with the lattice.

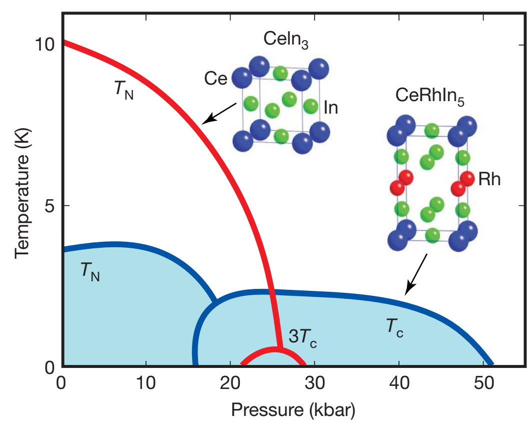


Figure 4 | Effect of electronic anisotropy. The schematic temperature–pressure phase diagram of two related heavy fermion compounds, $CeIn_3$ (refs 30, 52) and $CeRhIn_5$ (refs 53–55). These two materials differ in particular in the degree of anisotropy of the low energy excitation spectrum. As one would expect, the thermal fluctuations in the local magnetization lead to a smaller value of the magnetic transition temperature (Néel temperature, T_N) in the anisotropic material. By contrast, perhaps unexpectedly, T_c is greatly suppressed in the isotropic compound $CeIn_3$ (red lines) compared with $CeRhIn_5$ (blue lines). Both of these features are in qualitative agreement with the magnetic interaction model.

compound, $CeNi_2Ge_2$ (refs 57–59). The appearance of two superconducting domes in the temperature–pressure phase diagram of $CeNi_2Ge_2$ (refs 58, 59) led some of us to suspect a different origin of superconductivity in the two cases, that is, magnetically mediated on the border of antiferromagnetism at low pressures and density-mediated on the border of an α–γ density instability at higher pressures. Such magnetic and density instabilities have been identified in $CeCu_2Si_2$, in which the two superconducting domes coalesce in stoichiometric samples but become separate and distinct in Ge doped samples[60,61] (Fig. 5). Our model quasiparticle interaction can naturally account for this behaviour.

In our model, the overall scale for T_c depends on the characteristic width of the energy band out of which the paired quasiparticles emerge. This suggests that a further increase in T_c could be obtained if the 4*f* Ce band were replaced, for example, by a wider 5*f* band under otherwise equivalent conditions, for example, in the same tetragonal structure and without moving too far from the border of the magnetic instability. This is indeed what happens when Ce in $CeMIn_5$ is replaced by the 5*f* metal Pu to obtain the related systems $PuMGa_5$, which leads to an increase of T_c by about another order of magnitude[62,63].

These aspects of our model can be understood even with a simple Lindhard function that represents real-space oscillations with a single wavelength of the order of the lattice spacing, as is to be expected for a system on the border of commensurate antiferromagnetism. A whole new level of complexity and subtlety arises when Lindhard functions for realistic energy bands are used (Fig. 6). The oscillation pattern in real space is a superposition of components of different wavelengths, and can vary dramatically with the filling level of the relevant energy band.

These rich oscillation patterns have proved to be a critical element in developing a qualitative understanding of the scarcity of examples of superconductivity on the border of ferromagnetism[51]. Another crucial factor in the case of the border of ferromagnetism is that the magnitude of the quantum mechanical average of the inner product of the two quasiparticle spins is a factor of three smaller in the spin-triplet state than in the spin-singlet state. This unique feature of the magnetic interaction means that pairing in the spin-triplet state has an inherent disadvantage relative to pairing in the spin-singlet state. The factor of three in the relative ineffectiveness of the former pairing is a property of a quantum spin 1/2, and does not arise for a classical spin or in an anisotropic system in which only fluctuations in the magnetization along a particular axis in the crystal are relevant. This suggests that pairing on the border of ferromagnetism is more likely to arise in systems with a strong magnetic anisotropy[49]. This prediction motivated the search for superconductivity on the border of ferromagnetism in UGe_2, which is known to have a strong uniaxial magnetic anisotropy. The search proved fruitful, for it led to the discovery of the first example of the co-existence of superconductivity and itinerant-electron ferromagnetism[27,64–67] (Fig. 7). Soon after, superconductivity was discovered in the itinerant-electron ferromagnets URhGe (ref. 68) and UIr (ref. 69) and even in the ε-phase of elemental Fe (ref. 70).

The magnetic energy scales in *d* metals may be much larger than in the *f* metals and one might expect, on the basis of the magnetic interaction model, that the spin-triplet *p*-wave superconductivity in the nearly ferromagnetic tetragonal compound Sr_2RuO_4 would occur at relatively high temperatures, despite the inherent disadvantage of spin-triplet pairing discussed above. That this is not the case may be understood in terms of the subtle oscillation pattern of the Lindhard function for a realistic model of the 'd_{xy}' energy band of Sr_2RuO_4 thought to be dominant for superconductivity. In particular, the form of the Fermi surface that arises in the presence of a periodic lattice potential tends to have nesting features that lead to real-space oscillations in the magnetic interaction that interfere with the long-wavelength components associated with ferromagnetic correlations. An additional detrimental factor to *p*-wave pairing in Sr_2RuO_4 arises from the fact that strong ferromagnetic correlations only occur when the Fermi surface is close to a 'van Hove' singularity point. The Fourier transform of the *p*-wave Cooper state vanishes at these points, and hence the superconducting energy gap is small in regions of high density of electronic states on the Fermi surface—that

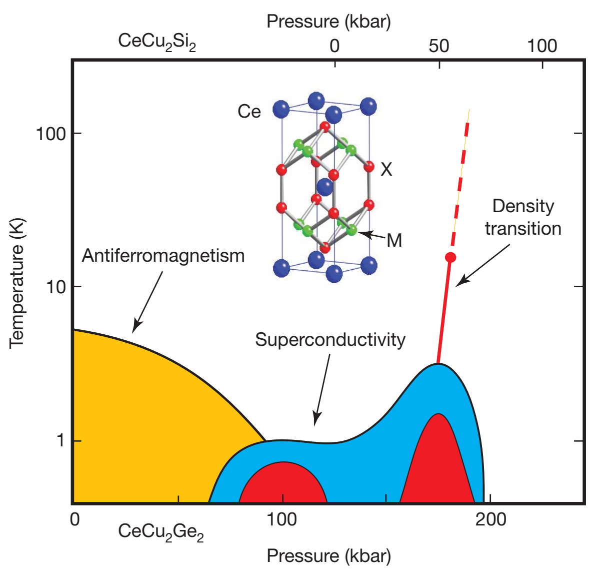


Figure 5 | Proximity to antiferromagnetic and density instabilities. Schematic temperature–pressure phase diagram for the class of compounds CeM_2X_2 having the crystal structure shown in the inset. In $CePd_2Si_2$ (ref. 30; that is, M = Pd, X = Si) and $CeRh_2Si_2$ (ref. 32), one only observes a superconducting dome on the border of metallic antiferromagnetism (left hand side of diagram), whereas in $CeNi_2Ge_2$ (refs 57–59) and in the close relatives $CeCu_2Si_2$ and $CeCu_2Ge_2$ (refs 4, 60, 61) one observes two superconducting domes, one of which is similar to that found in $CePd_2Si_2$ and $CeRh_2Si_2$. In $CeCu_2Si_2$ and $CeCu_2Ge_2$, the second superconducting dome (on the right hand side of the diagram) is found to be on the border of a density transition characterized by a first order transition line and low temperature critical end point (red dot) followed by crossover behaviour (red dashed line). This is analogous to a liquid–gas transition. (The two domes overlap in the pure compounds but clearly separate in chemically doped specimens (red domes).)

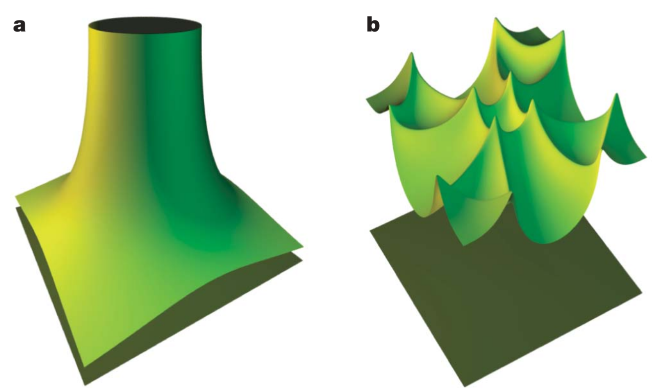


Figure 6 | Effect of crystal lattice on the magnetic properties. a, b, Static linear magnetic response (Lindhard function) versus wavevector in two dimensions for non-interacting electrons in a uniform background of neutralizing positive charges ('jellium'; **a**) and in the presence of a model crystal potential in a square lattice (**b**). Note the richer structure in the latter case. The different peaks give rise to oscillations of different wavelengths in real space, which can interfere constructively or destructively and thus lead to complex forms of the real-space magnetic interaction. As the Lindhard function is determined by the Fermi surface, it is as distinctive a property of a metal as is the Fermi surface itself.

is, p-wave pairing does not lead to an efficient lowering of the total electronic energy[50,51]. These and other factors discussed elsewhere shed light on why the search for a superconducting analogue of liquid ^{3}He has proved so elusive[27,50,51].

In the light of the experimental and theoretical findings discussed above, we can establish a number of conditions that are particularly favourable for magnetic pairing. On the basis of the cases studied so far, one class of promising candidate material should have, at minimum, the following properties: (1) a single band of relatively high characteristic energy scale; (2) strong quasi-two-dimensional antiferromagnetic correlations for spin-singlet pairing and for large amplitude oscillations of the spin–spin interaction; and (3) a crystal structure that enables the repulsive regions of the pairing potential to be optimally neutralized. The quasi-two-dimensional organic superconductors[28] appear to go quite far in satisfying these conditions.

Although the applicability of the magnetic interaction model to the copper oxides[71] has been disputed, it is nevertheless interesting that the model provides a natural explanation for the robustness of superconductivity and correctly anticipated the d-wave symmetry[34] of the Cooper-pair state in this class of materials[72–75], which appear to optimally satisfy all of the above conditions.

In the framework presented in this Review, the antiferromagnetic correlations come about through the Lindhard function—that is, the response of the non-interacting itinerant electron system to a space- and time-dependent applied magnetic field—but, crucially, as enhanced by an appropriate effective (molecular) field. This approach is most appropriate to the description of delocalized quasiparticles not too close to a metal–insulator transition. Another approach considers localized models of magnetism in which antiferromagnetic correlations can arise and also lead to d-wave superconductivity (see refs 76–80 for recent discussions). These two approaches are normally associated with itinerant versus localized models of magnetism, and are most naturally suited to the 'over-doped' and 'under-doped' states of the copper oxides, respectively. The detailed understanding of the similarities and differences of these approaches to magnetically mediated superconductivity would be a major step forward.

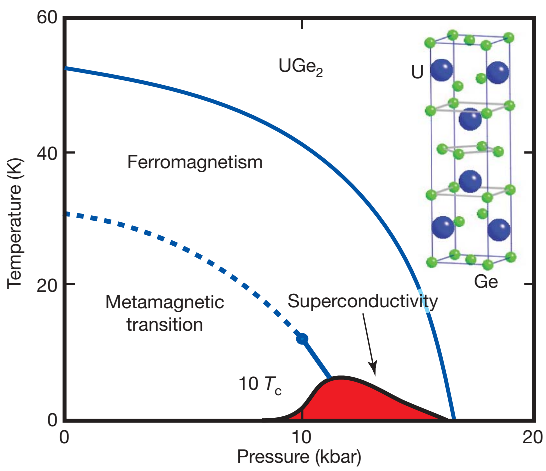


Figure 7 | Proximity to a ferromagnetic instability. Temperature–pressure phase diagram of the ferromagnetic superconductor UGe_2 (refs 27, 64, 65). The upper blue line corresponds to a ferromagnetic instability (analogous to the antiferromagnetic instability in Fig. 4) and the lower blue line to a metamagnetic instability with low critical end point (analogous to the density transition in Fig. 4). In this system, there is only one superconducting dome on the border of metamagnetism (a sudden change in the magnitude of the magnetization and density). A second superconducting dome is not observed on the border of the ferromagnetic transition. The strong first order nature of this transition, in contrast to the metamagnetic transition, leads one to expect relatively weak magnetic fluctuations and hence a weak magnetic pairing interaction.

The possibilities seem endless

In electrodynamics, the dimensionless coupling constant, which is a measure of the relative strength of the interaction, is 1/137 and hence small, leading only to minor corrections to the polarizer–analyser framework of the Coulomb interaction described earlier. The corresponding relevant dimensionless coupling constants characterizing the strength of the induced interaction (equation (2)) are likely to be much larger and possibly of order unity in real materials of interest here. It would therefore not be surprising if the lowest order interaction (equation (2)) fails to capture some essential physics. What is surprising is that it has led to so many useful insights and provided guidance in the searches for novel phenomena, some of which have proved to be fruitful.

The model interaction (equation (2)) was derived on the assumption that the characteristic amplitudes of the molecular fields are small. However, this is an unnecessary restriction, as the average over a gaussian distributed molecular field can now be carried out numerically by well-established Monte Carlo sampling techniques. One finds that when the characteristic dimensionless coupling constant is of order unity or greater, the model interaction (equation (2)) indeed fails to capture essential physics[81,82].

The quasiparticle-pairing problem where equation (2) ceases to be valid has not been studied in detail. At the level of approximation represented by equation (2), there is no fundamental difference in the various interaction channels (for example, density versus magnetic) in that, given the right conditions, they could all lead to robust forms of pairing. An understanding of the conditions favourable for high-temperature superconductivity cannot be obtained without considering the corrections to the simple theory[81,82]. This raises the question of whether one particular channel of interaction may be singled out by these higher corrections. In the preliminary non-perturbative (Monte Carlo) calculations, one finds that the magnetic interaction is enhanced by the higher order corrections, while the density interaction is reduced at least for the range of parameters considered. In this context, we note that in quantum electrodynamics the analogous corrections lead to an enhancement of the magnetic moment of the electron and therefore an enhancement of its magnetic coupling to the Maxwell field. The smallness of the enhancement in this latter case is due to the very weak coupling of the Maxwell field to the electron, in contrast to the case of the magnetic molecular field for which the corresponding enhancement can be substantial. These and previously mentioned features point to the special nature of spin in quantum mechanics, and may distinguish and favour the magnetic interaction over other interaction channels in the strong coupling limit.

The above non-perturbative treatment represents only one of many possible extensions of the model discussed in this Review for superconductivity without phonons. The richness of behaviour at even the simplest level of approximation points to the utmost importance of tuning material properties, and leaves no doubt that many more surprises are in store.

1. Laughlin, R. B. & Pines, D. The theory of everything. *Proc. Natl Acad. Sci. USA* **97**, 28–31 (2000).
2. Bardeen, J., Cooper, L. N. & Schrieffer, J. R. Theory of superconductivity. *Phys. Rev.* **108**, 1175–1204 (1957).
3. Parks, R. D. (ed.) *Superconductivity* Vols 1 and 2 (Dekker, New York, 1969).
4. Steglich, F. *et al.* Superconductivity in the presence of strong Pauli paramagnetism: $CeCu_2Si_2$. *Phys. Rev. Lett.* **43**, 1892–1896 (1979).
5. Ott, H. R., Rudigier, H., Fisk, Z. & Smith, J. L. UBe_{13} – an unconventional actinide superconductor. *Phys. Rev. Lett.* **50**, 1595–1598 (1983).
6. Stewart, G. R. Heavy-fermion systems. *Rev. Mod. Phys.* **56**, 755–787 (1984).
7. Fisk, Z. *et al.* Heavy-electron metals: New highly correlated states of matter. *Science* **239**, 33–42 (1988).
8. Taillefer, L. & Lonzarich, G. G. Heavy-fermion quasiparticles in UPt_3. *Phys. Rev. Lett.* **60**, 1570–1573 (1988).
9. Onnes, H. K. On the sudden change in the rate at which the resistance of mercury disappears. *Commun. Phys. Lab. Univ. Leiden* **124c**, 1 (1911).
10. Hoddeson, L. John Bardeen and theory of superconductivity. *J. Stat. Phys* **103**, 625–640 (2001).

11. Heisenberg, W. Zur Theorie der Supraleitung [On the theory of superconductivity]. *Z. Naturforsch. A* **2,** 185–201 (1947).
12. Macke, W. Uber die Wechselwirkungen im Fermi Gas-Polarisationserscheinungen, Correlationsenergie, Elektronenkondensation [On exchange effects in the Fermi gas-polarization features, correlation energy and electron condensation]. *Z. Naturforsch. A* **5,** 192–208 (1950).
13. Maxwell, E. Isotope effect in superconductivity of mercury. *Phys. Rev.* **78,** 477 (1950).
14. Reynolds, C. *et al.* Superconductivity of isotopes of mercury. *Phys. Rev.* **78,** 487 (1950).
15. Fröhlich, H. Theory of the superconducting state. I. The ground state at the absolute zero of temperature. *Phys. Rev.* **79,** 845–856 (1950).
16. Bardeen, J. Wave functions for superconducting electrons. *Phys. Rev.* **80,** 567–574 (1950).
17. Fröhlich, H. Interaction of electrons with lattice vibrations. *Proc. R. Soc. Lond. A* **215,** 291–298 (1952).
18. Bardeen, J. & Pines, D. Electron-phonon interaction in metals. *Phys. Rev.* **99,** 1140–1150 (1955).
19. Bohm, D. & Pines, D. A collective description of electron interactions: III. Coulomb interactions in a degenerate electron gas. *Phys. Rev.* **92,** 609–625 (1953).
20. Cooper, L. N. Bound electron pairs in a degenerate Fermi gas. *Phys. Rev.* **104,** 1189–1190 (1956).
21. Bednorz, J. G. & Müller, K. A. Possible high T_c superconductivity in the Ba-La-Cu-O system. *Z. Phys. B* **64** (2), 189–193 (1986).
22. Berk, N. F. & Schrieffer, J. R. Effect of ferromagnetic spin correlations on superconductivity. *Phys. Rev. Lett.* **17,** 433–435 (1966).
23. Nakajima, S. Paramagnon effect on the BCS transition in ^{3}He. *Prog. Theor. Phys.* **50,** 1101–1109 (1973).
24. Brinkman, W. F., Serene, J. W. & Anderson, P. W. Spin-fluctuation stabilization of anisotropic superfluid states. *Phys. Rev. A* **10,** 2386–2394 (1974).
25. Fay, D. & Appel, J. Coexistence of p-state superconductivity and itinerant ferromagnetism. *Phys. Rev. B* **22,** 3173–3182 (1980).
26. Maeno, Y. *et al.* Superconductivity in a layered perovskite without copper. *Nature* **372,** 532–534 (1994).
27. Saxena, S. S. *et al.* Superconductivity on the border of itinerant-electron ferromagnetism in UGe_2. *Nature* **406,** 587–592 (2000).
28. Jerome, D. *Organic Conductors* (Dekker, New York, 1994).
29. Fisk, Z. & Maple, M. B. On the existence of heavy fermion ytterbium compounds. *J. Alloy. Comp.* **183,** 303–311 (1992).
30. Mathur, N. D. *et al.* Magnetically mediated superconductivity in heavy fermion compounds. *Nature* **394,** 39–43 (1998).
31. Thalmeier, P., Zwicknagl, G., Stockert, O., Spain, G. & Steglich, F. in *Frontiers in Superconducting Materials* (ed. Narlikar, A.) 109–171 (Springer, Berlin, 2004).
32. Onuki, Y. *et al.* Recent advances in the magnetism and superconductivity of heavy fermion systems. *J. Phys. Soc. Jpn* **73,** 769–787 (2004).
33. Emery, V. J. The mechanisms of organic superconductivity. *Synth. Met.* **13,** 21–27 (1986).
34. Scalapino, D. J., Loh, E. & Hirsch, J. E. *d*-wave pairing near a spin-density wave instability. *Phys. Rev. B* **34,** 8190–8192 (1986).
35. Miyake, K., Schmitt-Rink, S. & Varma, C. M. Spin-fluctuation mediated even-parity pairing in heavy-fermion superconductors. *Phys. Rev. B* **34,** 6554–6556 (1986).
36. Landau, L. D. & Lifshitz, E. M. *Statistical Physics* 3rd edn, Pt 1 (Pergamon, Oxford, 1976).
37. Pines, D. & Nozières, P. *The Theory of Quantum Liquids – Normal Fermi Liquids* (Addison-Wesley, Wokingham, 1989).
38. Aldrich, C. H. & Pines, D. Polarisation potentials and elementary excitations in liquid-He-3. *J. Low Temp. Phys.* **32,** 689–715 (1978).
39. Kohn, W. & Luttinger, J. M. New mechanism for superconductivity. *Phys. Rev. Lett.* **15,** 524–526 (1965).
40. Moriya, T. *Spin Fluctuations in Itinerant Electron Magnetism* (Springer, Berlin, 1985).
41. Millis, A. J. Effect of a non-zero temperature on quantum critical points in itinerant fermion systems. *Phys. Rev. B* **48,** 7183–7196 (1993).
42. Lonzarich, G. G. in *Electron* (ed. Springford, M.) Ch 6 (Cambridge Univ. Press, Cambridge, UK, 1997).
43. Monthoux, P. & Lonzarich, G. G. Density-fluctuation-mediated superconductivity. *Phys. Rev. B* **69,** 064517 (2004).
44. Osheroff, D. D., Richardson, R. C. & Lee, D. M. Evidence for a new phase of solid He-3. *Phys. Rev. Lett.* **28,** 885–888 (1972).
45. Monthoux, P. & Lonzarich, G. G. Magnetically mediated superconductivity in quasi-two and three dimensions. *Phys. Rev. B* 63, 054529 (2001); Magnetically mediated superconductivity: Crossover from cubic to tetragonal lattice. *Phys. Rev. B* **66,** 224504 (2002).
46. Arita, R., Kuroki, K. & Aoki, H. *d*- and *p*-wave superconductivity mediated by spin fluctuations in two-and three-dimensional single-band repulsive Hubbard model. *J. Phys. Soc. Jpn* **69,** 1181–1191 (2000).
47. Fukazawa, H. & Yamada, K. Theory on superconductivity of $CeIn_3$ in heavy fermion systems. *J. Phys. Soc. Jpn* **72,** 2449–2452 (2003).
48. Takimoto, T., Hotta, T. & Ueda, K. Strong-coupling theory of superconductivity in a degenerate Hubbard model. *Phys. Rev. B* **69,** 104504 (2004).
49. Monthoux, P. & Lonzarich, G. G. *p*-wave and *d*-wave superconductivity in quasi-two-dimensional metals. *Phys. Rev. B* **59,** 14598–14605 (1999).
50. Honerkamp, C. & Rice, T. M. Cuprates and ruthenates: Similarities and differences. *J. Low Temp. Phys.* **131,** 159–167 (2003).
51. Monthoux, P. & Lonzarich, G. G. Magnetic interactions in a single-band model for the cuprates and ruthenates. *Phys. Rev. B* **71,** 054504 (2005).
52. Knebel, G., Braithwaite, D., Canfield, P. C., Lapertot, G. & Flouquet, J. The quantum critical point revisited in $CeIn_3$. *High Press. Res.* **22,** 167–170 (2002).
53. Hegger, H. *et al.* Pressure-induced superconductivity in quasi-2D $CeRhIn_5$. *Phys. Rev. Lett.* **84,** 4986–4989 (2000).
54. Petrovic, C. *et al.* Heavy-fermion superconductivity in $CeCoIn_5$ at 2.3 K. *J. Phys. Condens. Matter* **13,** L337–L342 (2001).
55. Thompson, J. D. *et al.* Magnetism and unconventional superconductivity in $Ce_nM_mIn_{3n+2m}$ heavy-fermion crystals. *Physica B* **329–333,** 446–449 (2003).
56. Curro, N., Fisk, Z. & Pines, D. Scaling and the magnetic origin of emergent behaviour in correlated electron superconductors. *MRS Bull.* **30,** 442–446 (2005).
57. Gegenwart, P. *et al.* Non-Fermi liquid effects at ambient pressure in stoichiometric heavy-fermion compound with very low disorder: $CeNi_2Ge_2$. *Phys. Rev. Lett.* **82,** 1293–1296 (1999).
58. Grosche, F. M. *et al.* Anomalous low temperature states in $CeNi_2Ge_2$ and $CePd_2Si_2$. *J. Phys. Condens. Matter* **12,** L533–L540 (2000).
59. Braithwaite, D. *et al.* Superconductivity, upper critical field and normal state resistivity in $CeNi_2Ge_2$ under pressure. *J. Phys. Condens. Matter* **12,** 1339–1349 (2000).
60. Yuan, H. Q. *et al.* Observation of two distinct superconducting phases in $CeCu_2Si_2$. *Science* **302,** 2104–2107 (2003).
61. Holmes, A. T., Jaccard, D. & Miyake, K. Signatures of valence fluctuations in $CeCu_2Si_2$ under high pressure. *Phys. Rev. B* **69,** 024508 (2004).
62. Sarrao, J. L. *et al.* Plutonium-based superconductivity with a transition temperature above 18 K. *Nature* **420,** 297–299 (2002).
63. Bauer, E. D. *et al.* Structural tuning of unconventional superconductivity in $PuMGa_5$ (M = Co, Rh). *Phys. Rev. Lett.* **93,** 147005 (2004).
64. Huxley, A. *et al.* The co-existence of superconductivity and ferromagnetism in actinide compounds. *J. Phys. Condens. Matter* **15,** S1945–S1955 (2003).
65. Kitaoka, Y. *et al.* Novel phase diagram of superconductivity and ferromagnetism in UGe_2: A ^{73}Ge-NQR study under high pressure. *J. Phys. Condens. Matter* **17,** S975–S986 (2005).
66. Coleman, P. Superconductivity – On the verge of magnetism. *Nature* 406, 580–581 (2000); Superconductivity – Magnetic glue exposed. *Nature* **410,** 320–321 (2001).
67. Flouquet, J. & Buzdin, A. Ferromagnetic superconductors. *Phys. World* **15,** 41–46 (2002).
68. Aoki, D. *et al.* Coexistence of superconductivity and ferromagnetism in URhGe. *Nature* **413,** 613–616 (2001).
69. Akazawa, T. *et al.* Pressure-induced superconductivity in UIr. *J. Phys. Soc. Jpn* **73,** 3129–3134 (2004).
70. Shimizu, K. *et al.* Superconductivity in the nonmagnetic state of iron under pressure. *Nature* **412,** 316–318 (2001).
71. Pines, D. Spin excitations and superconductivity in cuprate oxide and heavy electron superconductors. *Physica B* **163,** 78–88 (1990).
72. Moriya, T., Takahashi, Y. & Ueda, K. Antiferromagnetic spin fluctuations and superconductivity in two-dimensional metals — a possible model for high T_c oxides. *J. Phys. Soc. Jpn* **59,** 2905–2915 (1990).
73. Monthoux, P. *et al.* Toward a theory of high-temperature superconductivity in the antiferromagnetically correlated cuprate oxides. *Phys. Rev. Lett.* **67,** 3448–3451 (1991).
74. Wollman, D. A. *et al.* Experimental determination of the superconducting pairing state in YBCO from the phase coherence of YBCO-Pb DC SQUIDs. *Phys. Rev. Lett.* **71,** 2134–2137 (1993).
75. Tsuei, C. C. *et al.* Pairing symmetry and flux quantization in a tricrystal superconducting ring of $YBA_2Cu_3O_{7-\delta}$. *Phys. Rev. Lett.* **73,** 593–596 (1994).
76. Anderson, P. W. *et al.* The physics behind high-temperature superconducting cuprates: The 'plain vanilla' version of RVB. *J. Phys. Condens. Matter* **16,** R755–R769 (2004).
77. Lee, P. A., Nagaosa, N. & Wen, X.-G. Doping a Mott insulator: Physics of high temperature superconductivity. Preprint at ⟨http://arxiv.org/abs/cond-mat/0410445v1⟩ (2004).
78. Arrigoni, E., Fradkin, E. & Kivelson, S. A. Mechanism of high-temperature superconductivity in a striped Hubbard model. *Phys. Rev. B* **69,** 214519 (2004).
79. Baskaran, G. Mott insulator to high T_c superconductor via pressure: Resonating valence bond theory and prediction of new systems. *Phys. Rev. Lett.* **90** (19), 197007 (2003).
80. Barzykin, V. & Pines, D. Phenomenological model of protected behaviour in the pseudogap state of underdoped cuprate superconductors. *Phys. Rev. Lett.* **96,** 247002 (2006).
81. Monthoux, P. Migdal's theorem and the pseudogap. *Phys. Rev. B* **68,** 064408 (2003).
82. Monthoux, P. Diagrammatic perturbation theory and the pseudogap. *Phys. Rev. B* **70,** 144403 (2004).

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to G.G.L. (GGL1@phy.cam.ac.uk).

REVIEWS

After a dozen years of progress the origin of angiosperms is still a great mystery

Michael W. Frohlich[1] & Mark W. Chase[1]

Here we discuss recent advances surrounding the origin of angiosperms. Putatively primitive characters are now much better understood because of a vastly improved understanding of angiosperm phylogenetics, and recent discoveries of fossil flowers have provided an increasingly detailed picture of early diversity in the angiosperms. The 'anthophyte theory', the dominant concept of the 1980s and 1990s, has been eclipsed; Gnetales, previously thought to be closest to the angiosperms, are related instead to other extant gymnosperms, probably most closely to conifers. Finally, new theories of flower origins have been proposed based on gene function, duplication and loss, as well as on morphology. Further studies of genetic mechanisms that control reproductive development in seed plants provide a most promising avenue for further research, including tests of these recent theories. Identification of fossils with morphologies that convincingly place them close to angiosperms could still revolutionize understanding of angiosperm origins.

Less than a dozen years ago even the most basic questions regarding the origin of angiosperms were still disputed, including the nature of primitive flowers, what sorts of gymnosperms might have given rise to angiosperms, and the broadest outlines of the evolutionary trajectory between them. (See ref. 1 for references not cited here.) Studies of fossil flowers[2] showed that bisexual and unisexual flowers both occurred in the earliest fossil flower floras, so it was still possible that plants with unisexual flowers consisting of a single stamen or a single carpel (resembling extant *Ceratophyllum* or *Hedyosmum*) might reflect the ancestral angiosperm condition. Other analyses supported the directly opposite view: the overall organization of the bisexual angiosperm flower (flat structures surrounding male organs surrounding central female organs) had been inherited directly from gymnosperm ancestors. Subsequent advances, derived from new data and reinterpretations of older data, have narrowed the range of alternative explanations for origins of both flowers and angiosperms. New data have come especially from molecular phylogenetics, but also from studies of gene function, duplication and loss and from palaeobotany. The rise of evolutionary developmental biology ('evo–devo') has reinvigorated the study of plant anatomy and led to new, increasingly synthetic theories; they seek to fuse disparate fields to explain various aspects of flower origins. Formulation of detailed, testable theories combined with study of fossils and genes has the power to dispel the mystery surrounding the origin of both flowers and angiosperms.

Relationships between extant angiosperms

Relationships of relatively few groups of angiosperms are still in dispute. Furthermore, morphological and molecular phylogenetic results are now considerably more congruent than in the past. Analyses in the past five years have consistently pointed to the 'ANA' (formerly called ANITA) taxa—Amborellaceae, Nymphaeales and Austrobaileyales—as successive sister groups to the larger clades of magnoliids, eudicots and monocots (Fig. 1), although there has been some dispute over the relative positions of Amborellaceae and Nymphaeales[3]. A noteworthy new addition near the basal nodes of the angiosperm tree is the Hydatellaceae, a small family of minute aquatics with small simple flowers that were previously thought to be members of the monocot order Poales. These fall as sister to Nymphaeales (Fig. 1) and extend the range of morphologies among these clades[4,5].

The ANA taxa, including Hydatellaceae, are each individually highly specialized. For example, *Amborella* grows in wet, forest understorey habitats in New Caledonia and is dioecious (but with vestigial organs of the opposite sex), whereas Nymphaeales with perfect flowers are all adapted to aquatic habitats. Improvements in morphological reconstructions of primitive angiosperms can yet be expected (particularly with the application of likelihood methods that consider branch lengths in projecting character states down to the basal node[1]), but the range of hypotheses now considered relevant is considerably narrower than in the past. For example, the old view of the primitive carpel as conduplicate—folded lengthways and fused at the edge—was based on magnoliid taxa now known to be relatively derived. Carpels of most ANA taxa are bucket-shaped and sealed only by a secretion[3]. Most importantly, several previously popular ideas can now be discarded, such as the idea that angiosperms arose from more than one 'gymnosperm' ancestor.

Fossil flowers

The only direct evidence of early angiosperm flowers comes from fossils. Mesofossils, up to a few millimetres in size, often show exquisite three-dimensional preservation, including cell structure. Most mesozoic flowers fall within the mesofossil range. Diverse mesofossil assemblages span the late-Early to Late Cretaceous period (reviewed in ref. 6), and others extending back to the upper Jurassic are now known (E.-M. Friis, personal communication). Flowers referable to ANA angiosperms are found in the earliest mesofossil assemblages, along with flowers of Chloranthaceae, which are sister to the magnoliids in recent analyses of complete sets of genes from the plastid genome[53]. Fossil pollen provides yet earlier evidence of angiosperms at roughly 136 Myr ago (Hauterivian[6]; mid-Early Cretaceous), about 10 Myr before the earliest published mesofossil floras[7]. Fossil pollen

[1]Royal Botanic Gardens Kew, Richmond, Surrey TW9 3DS, UK.

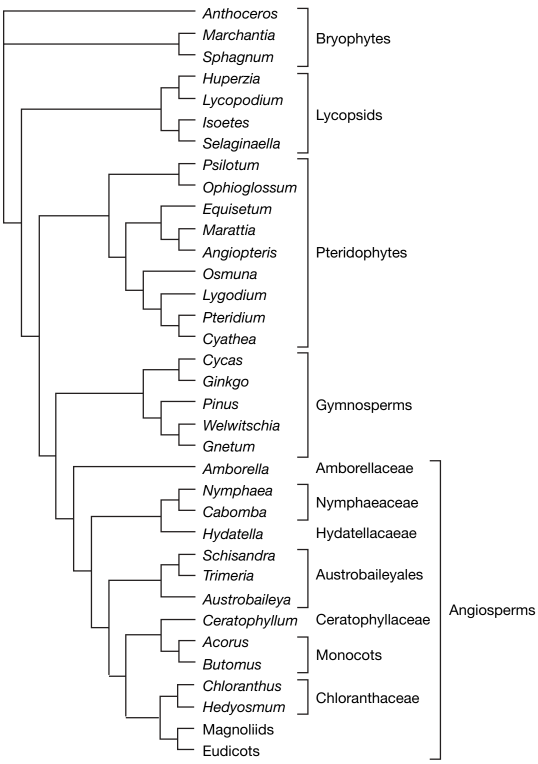


Figure 1 | The phylogenetic tree of living plant groups (based on refs 3, 51).

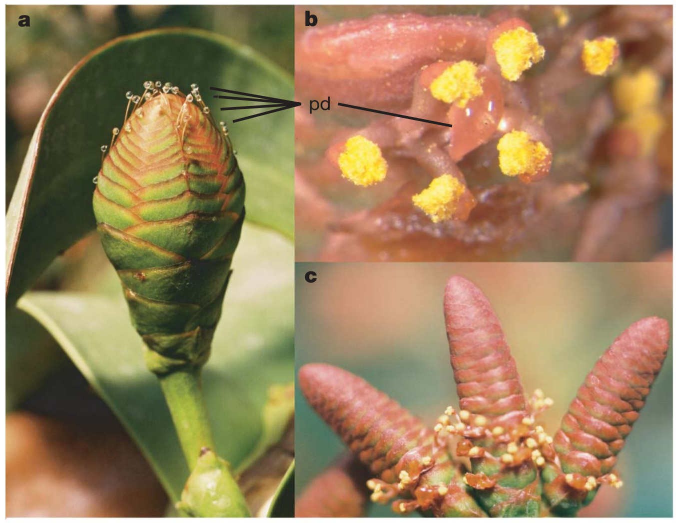


Figure 2 | *Welwitschia* cones. **a**, Female. **b**, Close-up of male cone, showing pollen organs and pollination droplet in between them. **c**, Male cones. pd, pollination droplet.

shows that diverse magnoliids, monocots and early eudicots had appeared by the early Aptian, about 125 Myr ago, demonstrating an early, rapid major radiation.

Molecular results have sharpened evolutionary interpretations. For example, unisexual fossil flowers similar to the extant genus *Hedyosmum* (Chloranthaceae) are among the earliest flowers known[6]. However, our knowledge that Chloranthaceae insert well above the ANA taxa clearly indicates that unisexuality in *Hedyosmum* and most probably also these fossil taxa is due to secondary reduction. The addition of Hydatellaceae as sister to Nymphaeales does not change this inference.

There are no studied fossils clearly representing stem-group angiosperms, that is, of plants related to extant angiosperms but attached below the basal node of extant angiosperms in the tree. Such fossils might provide spectacular direct evidence of morphological change along this unknown stretch of evolutionary history. *Archaefructus*, originally thought to be a stem-group angiosperm of Jurassic age, is not; it has been re-dated as mid-Early Cretaceous, and its reproductive unit has been reinterpreted as an inflorescence, not a flower[8,9]. Reinterpretation of *Archaefructus* is a good example of initial morphological interpretations leading to remarkably different ideas of relationship compared with subsequent analyses.

Evidence from phylogenetics and morphology

Living gymnosperms and angiosperms constitute the extant seed plants. The four groups of living gymnosperms are only a remnant of the substantial diversity of Palaeozoic and Mesozoic times[10]. In the 1980s, morphological cladistic analyses of living and fossil seed plants[11,12] placed cycads sister to the other living taxa and identified 'anthophytes' as consisting of angiosperms, Gnetales (Fig. 2) and two extinct groups, Bennettitales (Fig. 3e) and *Pentoxylon*. The first three include members with reproductive units that have sterile appendages surrounding male structures with female structures in the centre, suggesting that this overall organization might be homologous in the three groups, hence much antedating origin of angiosperms. Otherwise, their reproductive structures differ markedly, but, given these relationships, morphological transformations have been proposed[12]. Subsequent analyses placed Caytoniales (Fig. 3b, c), which lack this overall organization, within anthophytes as sister to angiosperms, undermining this supposed homology in overall organization[13].

Molecular phylogenetic analyses of seed plants now indicate that living gymnosperms are monophyletic, with Gnetales related to conifers[1,14], although this remains controversial[6,15,16]. Palaeobotanists are increasingly willing to consider extant gymnosperm monophyly, but with varying levels of surprise and disquiet over the implications[17]. In the two most recent morphological analyses, placing Gnetales with conifers made trees one step longer[17] or forcing extant gymnosperm monophyly cost four additional steps[15], showing that the signal against extant gymnosperm monophyly is not especially strong.

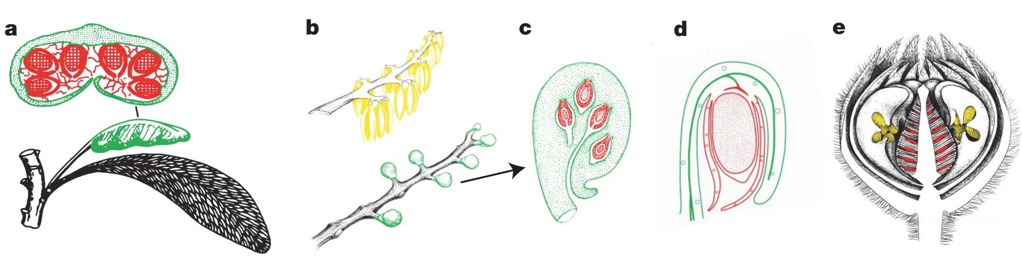


Figure 3 | Fossil gymnosperms. **a**, *Glossopteris* showing cupules borne on stalk above a leaf (from ref. 23). **b**, *Caytonia* male (above) and female (below) reproductive units (from ref. 11). **c**, *Caytonia* cupule (from ref. 36). **d**, Corystosperm (*Umkomasia*) cupule containing one ovule (from ref. 52). Cupule wall almost surrounds ovule, except for a slit facing the stalk. **e**, Bennettitales (*Williamsoniella*) bisexual reproductive unit (from ref. 11); each oval pollen sac consists of several fused microsporangia. Ovules are borne among scales on the central stalk; in *Vardekloftia* each is enclosed by a cupule wall. Green, cupule wall; red, ovule; yellow, pollen organ.

Extant gymnosperm monophyly moves Gnetales and cycads over many nodes compared with the shortest morphology-based trees[15,17], but only rarely have phylogenetics studies of morphology and DNA data agreed in plant studies, even in well-studied groups. Early morphological cladistic analyses of angiosperms underwent a radical rethinking of character homologies in the light of DNA analyses, which generated results much more in line with the DNA trees[18]. Interpretation of morphological homologies can radically shift if evidence of alternative relationships triggers re-examination; however, no source of phylogenetic evidence is infallible. Future studies should show whether current molecular or morphological results are erroneous.

New types of data are promising, from large-scale sequencing of nuclear genes[19] to molecular fossils. Oleanane is a diagenetic product of triterpenoids found in most angiosperms. Taylor *et al.*[20] demonstrated oleanane associated with several Bennettitales and Permian gigantopterid fossils, but oleananes were not found in Gnetales, Palaeozoic medullosan pteridosperms or in the conifer relatives Cordaitales. Other chemical fossils are markers for Cordaitales[21].

Monophyly of extant gymnosperms places them all equally distant from the angiosperms, which means that the lineage that eventually produced angiosperms diverged from the common ancestor with extant gymnosperms much earlier than previously thought, from among the 'pteridosperms' ('seed ferns'). Living gymnosperms show a great diversity of reproductive morphologies, and these must have resulted from numerous specializations. This makes comparison with angiosperms much more difficult.

Some extinct 'gymnosperm' groups must be closely related to angiosperms. If living and fossil 'gymnosperms' are considered together, then angiosperms arose from within them, making 'gymnosperms' paraphyletic (which we indicate with quotation marks). Both Caytoniales (Fig. 3b, c) and Bennettitales (Fig. 3e), fossil 'gymnosperms' with remarkably different morphologies, have long figured in theories of angiosperm origins[11,22] and appear as successive sister groups to angiosperms in recent studies[15,17]. Caytoniales have cupules that could plausibly be transformed into angiosperm bitegmic ovules (see Box 1), but these are borne on slender stalks unlike carpels. The detached male structures also differ significantly from angiosperm stamens, and it is not known how either of these was borne on the plant. Some Bennettitales have bisexual reproductive units[11], but there is no obvious carpel precursor, and except for *Vardekloeftia*[17] the ovules are not borne inside a cupule, so the source of angiosperm-type bitegmic ovules is also uncertain. Microsporophylls are highly variable, but none so far known closely resemble angiosperm stamens.

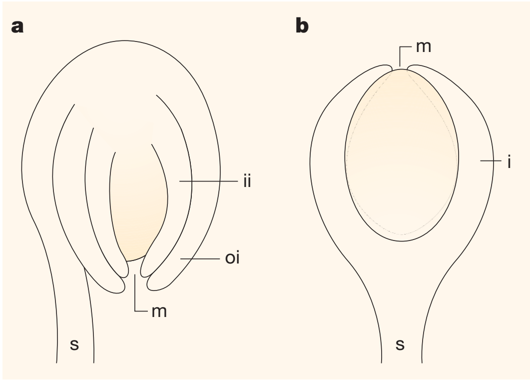
Retallack and Dilcher[23] suggested that the angiosperm carpel could be derived from structures resembling those of glossopterids (Fig. 3a), a group of Permian 'gymnosperms' that had a cupule or cupules borne on stalks above foliage leaves. Glossopterids (Fig. 3a) had one or more cupules borne on stalks above foliage leaves and have been suggested as angiosperm ancestors[23], but their early (Permian) age is problematic. Doyle[17] suggested that Caytoniales may be related to glossopterids and may also have had the reproductive stalk borne above a subtending leaf. If the stalk became fused to the leaf, the resulting structure would be an ideal carpel precursor.

Friedman and Floyd[24] proposed a theory about the angiosperm female gametophyte that uses the idea of developmental modules to account for arrangements and fates of nuclei, including those that participate in double fertilization to make zygote and endosperm (food store). They suggested that the basic module consists of four nuclei, one of which moves to the centre of the initially coenocytic gametophyte to fuse with the second sperm forming the endosperm nucleus. In the module near the micropyle, where the pollen tube enters, the other three nuclei organize the two synergids and the egg cell that fuses with a sperm, making the zygote. Most angiosperms have a second module that also sends a nucleus to the centre of the gametophyte, so fusion generates the standard, triploid endosperm nucleus. There is much variation on this basic pattern, but most of this diversity is explicable by changing the numbers of modules. Nymphaeaceae have only one module, which could be the ancestral condition, especially because *Amborella* has a unique system.

Some unreasonable theories posit multiple origins of angiosperms from 'gymnosperm' ancestors[25,26]. Angiosperms have many shared derived characters[11,12], and it is most unlikely that such complex features, arising independently, would fail to show differences that reveal their independent origins. All molecular and morphological analyses support angiosperm monophyly.

Box 1 | Ovules and cupules

In both gymnosperms and angiosperms, seeds develop from ovules. Ovules consist of a stalk that bears the nucellus (megasporangium) at its tip, with one or two covering layers (integuments) that grow from the stalk and almost completely cover the nucellus (Box 1 Figure). Only a small opening (micropyle) through the integuments allows access to the nucellus, in which the megagametophyte and egg(s) form.



Box 1 Figure | Diagrammatic ovules. a, Angiosperm ovule. **b**, Gymnosperm ovule. i, integument (covering); ii, inner integument; m, micropyle (opening); oi, outer integument; s, stalk. (From ref. 38.)

Gymnosperm ovules have a single integument (unitegmic) with the micropyle positioned opposite the stalk (although modified bracts in Gnetales have been called additional integuments). Most angiosperms, including all ANA taxa, have two integuments (bitegmic; reduced to one in some derived groups), and the ovule is commonly bent over, positioning the micropyle close to the stalk. The second integument is not simply a reiteration of the first; the two clearly have different attributes[22], which poses questions about the origin of the second (outer) integument.

Some fossil gymnosperms bear ovules inside an enclosing structure (cupule wall; Fig. 3) that is derived from a single modified leaf or part of a leaf. In Caytoniales and Corystospermales, ovules are almost completely enclosed, except for a small opening near the stalk of the cupule. If the cupule has only a single gymnosperm-style ovule inside it (as in some corystosperms; Fig. 3d), the whole structure resembles an angiosperm ovule, with the cupule wall comparable to the outer integument of a bitegmic angiosperm ovule and the single integument of the gymnosperm ovule comparable to the inner integument of the angiosperm ovule[22]. Glossopteridales also have cupules, although they usually do not tightly enclose the numerous ovules (Fig. 3a).

It is not clear whether cupules of all Mesozoic gymnosperms are homologous or whether there were multiple origins of cupules. They differ in the dorsiventrality of the cupule wall: in corystosperms the outer surface is comparable to the upper side of a leaf[52] (adaxial surface), on the basis of the arrangement of vascular tissues, whereas probably in Caytoniales and certainly in glossopterids (and in the poorly known *Petriellaea*) it is comparable to the lower side (abaxial surface). Some angiosperms develop vascular tissue in the outer integument, with an orientation suggesting that the outer surface is morphologically abaxial, pointing to Caytoniales or Glossopteridales as possible angiosperm antecedents[37].

MADS genes

MADS-box transcription factors are important for flower origins because they specify the major floral organs and because their expression zones typically correspond to their zones of action, so expression studies are useful for inferring function. According to the classic 'ABC' model (in *Arabidopsis* terminology), sepals are specified by the 'A' gene *Ap1* (and the non-MADS *AP2*), petals by the 'A' in combination with the two 'B' genes *PI* and *AP3*, stamens by the 'B' genes and the 'C' gene *AG* and carpels by *AG* alone. *SEPALLATA* (*SEP*) or 'E' genes are now known to be required for all four organ types, and the 'D' genes have been proposed as specifying ovules. Expression of the A, B, C and E MADS genes is upregulated by *LEAFY*, a non-MADS transcription factor.

Gene phylogenetics shows that each major MADS subgroup extends back to the base of extant angiosperms. There have been many duplication events within these clades, some probably reflecting whole-genome duplications, such as at the base of eudicots[19].

Several pairs of major MADS clades result from duplications below extant angiosperms, such as the *PI* and *AP3* clades, and also the *AG* clade and the putative 'D' gene clade. The most closely related gymnosperm genes are sister(s) to these clade pairs.

Classical 'A' function may be limited to relatives of *Arabidopsis*, whereas in other plants 'A' function may not be separable from the other major role of the 'A' genes in helping to specify apices as floral[27]. Lack of a unique sepal-specifying system is consistent with the suggestion that in the original flower the perianth may have been composed entirely of petals[28].

Understanding the specification of ovules would be especially important but remains problematic. Overexpression of a *Petunia* 'D' gene in *Petunia* generates ectopic ovules on sepals and petals, but overexpression of the *Arabidopsis* orthologue in *Arabidopsis* does little. It is unclear whether ovule function versus stamen + carpel function characterized the 'D' versus 'C' clades from their initial divergence. Kramer *et al.*[29] concluded that they do, but Zahn *et al.*[30] produced contradictory evidence. Specification of ovules and their components is highly complex[31], and there may be differences between taxa.

In ANA angiosperms, 'B' (and to some degree 'C') MADS genes show much broader messenger RNA expression than in eudicots, which has led to the 'fading borders' model of floral organ specification[19,32]. This posits an activity gradient of floral genes that determine organ identity, resulting in a gradient in organ morphology from the outside to the centre, in contrast to flowers of most eudicots, which have sharply distinguished organs. Many ANA grade flowers have variable numbers of floral organs often arranged spirally (instead of in whorls), suggesting less developmental homeostasis than in eudicots or monocots[1,19,32]. Perhaps lower developmental homeostasis implies simpler systems for specifying floral organs in these plants that are more appropriate for comparison with gymnosperms than are systems of other angiosperms.

In gymnosperms, 'B' gene homologues are primarily expressed in developing male structures, resembling their role in angiosperm stamens. Homologues of the 'C' + 'D' clade are expressed in both male and female structures (including ovules), also suggesting broad conservation in their roles. These apparently conserved functions allow theories of flower origin based on these genes[33–35].

Theories based on evo–devo analysis

The 'mostly male' theory[1,36–38] was triggered by studies of the *LEAFY* gene. It has two paralogues in gymnosperms but only one in angiosperms, in which it helps specify the flower. Data from pine suggested that the two gymnosperm paralogues may specify male versus female cones. Angiosperms have lost the latter copy, suggesting that the flower may be derived mostly from the male reproductive unit. At the extreme, the minimal female structure (for example a cupule) might have moved onto the male structure ectopically, creating the antecedent for the carpel bearing angiosperm-type ovules (Fig. 4).

This is supported by other observations: within carpels, ovules have highly variable numbers and placements, some of which must represent ectopic movement and increase in numbers. Stamens, by contrast, are highly uniform. *Arabidopsis* null mutants of *lfy* make no stamens but still form carpels, showing that *LFY* is required for male specification, but *LFY* independent genes can specify carpels. Ectopic ovules can be generated in *Petunia* by the overexpression of a single gene and in *Arabidopsis* by a different mutation, and functional ectopic ovules occur naturally on leaves of some plants of *Ginkgo* (a gymnosperm) (Fig. 5), suggesting that ectopic ovules are relatively easily produced. Liquid exuded by sterile ovules in Gnetales (Fig. 2) attracts insects to male structures, and pollination droplets on the functional female ovules also attract insects, resulting in pollination. Ectopic ovule placement in male cones in angiosperm ancestors might have conferred an immediate selective advantage by encouraging insect pollination[36].

Recent work on three conifers found complex patterns of expression of the two paralogues, with both being expressed at about equal levels in early female cone development, and both being expressed in early-developing male structures[39]. This argues against any role in specifying male versus female cones. This undermines the mostly male theory, but preliminary expression data from *Welwitschia* favours the theory (E. Moyroud and M.W.F., unpublished observations), and the supporting morphological evidence remains.

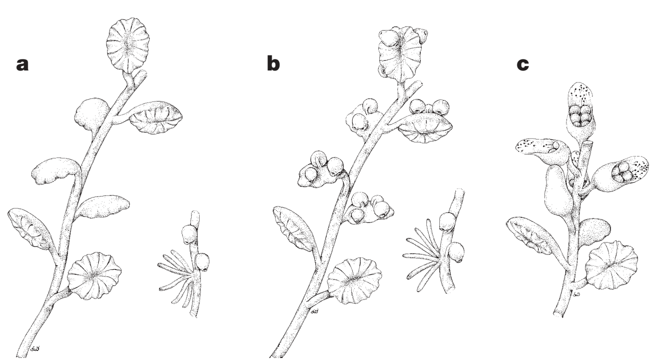


Figure 4 | Steps in the mostly male theory. a, Gymnosperm with separately borne microsporophylls (male; left) and cupules (female; right). **b**, Cupules have moved ectopically onto some microsporophylls. **c**, Microsporophylls bearing cupules are transforming into carpels and cupules into angiosperm-style ovules (from ref. 38).

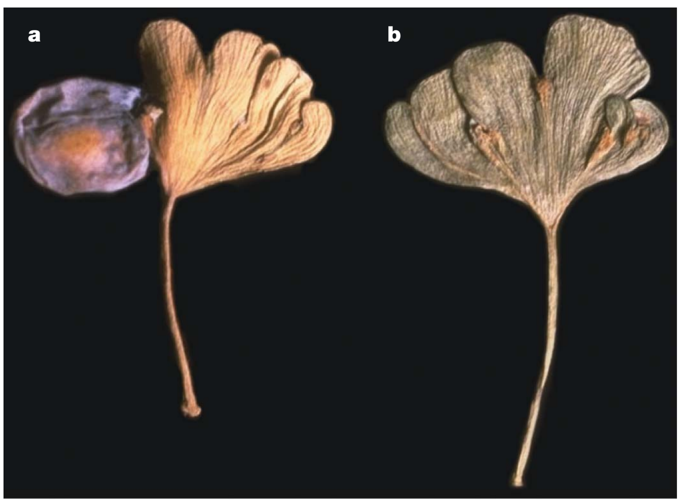


Figure 5 | *Ginkgo* leaves bearing ectopic ovules (and showing autumn colour). a, Leaf bearing mature seed at the base of an indentation. **b**, Leaf with many indentations that have ectopic ovules. (From ref. 37.)

An intriguing observation is the large number of mutations that can homeotically transform the outer integument of an ovule into a carpelloid structure[31], which is consistent with the homology of both of these to a leaf. It might also suggest some ancestral structure resembling cupules borne within a cupule, but no such structure is known among fossil gymnosperms. An alternative would be the spread of some elements of cupule-wall identity from ectopic cupules onto the microsporophyll that bears them, precipitating modification of the latter into a carpel wall.

The old 'gametoheterotopy' theory of Meyen[40] remains relevant. It begins with the bisexual reproductive structures of Bennettitales and posits a homeotic conversion that partly imposes morphology of the pollen-bearing organs onto female structures. Some Bennettitalian microsporophylls were flattened with pollen organs on their upper (adaxial) surface. If homeotic conversion resulted in ovules borne on upper surfaces of flat microsporophyll-like structures, then the combination would serve as a carpel precursor. Bennettitales typically have a single whorl of male structures, so homeotic transformation rather than ectopic placement of ovules onto pre-existing structures would be required.

The 'out-of-male/out-of-female' theory of Theissen *et al.*[33,34] centred on the origin of flower bisexuality. They noted that modern conifers sometimes make bisexual cones. Downregulation of 'B' gene expression in the distal portion of a male gymnosperm cone could permit the tip to become female, or upregulation of 'B' gene expression in basal regions of a female cone could make that region male, generating bisexual reproductive units from either male or female ancestral structures. They suggest that insect pollination could confer an immediate selective advantage, as in the mostly male theory. However, in conifers the resulting cones show normal male and female morphology in both cone regions. There is no novel morphology beyond bisexuality, so the origin of the angiosperm carpel structure is not explained.

Baum and Hileman[35] proposed a theory that adds mechanistic detail to the out-of-male theory. They suggested that greatly increased expression of protein encoded by a 'C' gene in the terminal region of a male cone could have been complexed with all the *sepallata*-encoded protein, preventing its interaction with the protein encoded by the 'B' gene, so switching its developmental fate to female. They also suggested that this 'C' protein might have repressed *WUSCHEL*, a gene required for maintenance of the apical meristem, resulting in floral determinacy.

The more explicit a theory is, the easier it is to test, so such explicit theories are especially valuable. These recent theories differ from earlier views in the crucial aspect of being testable, not only through the discovery of fossils but also by data from evo–devo studies.

The future

We certainly hope that spectacular palaeobotanical discoveries will clarify flower origins, but failing that it is evo–devo studies that will provide the most important new data, both by suggesting and testing theories of flower origin. Even simple gene-expression data may help in judging whether particular structures are homologous or not[41], although such comparisons can be misleading[42]. Vestigial genes or gene expression patterns may indicate directions of evolutionary change[43]. Gene trees analysed within organismal trees offer special power for detecting neofunctionalization as opposed to retained (plesiomorphic) gene functions[44]. All of these results help to limit the range of possible theorizing, which we hope will converge on a historically accurate account of flower origins.

The comparative method applied to morphology and development fuelled the first great advances in evolutionary understanding, and similarly the comparative method applied to DNA sequences revolutionized our understanding of phylogenetic relationships between land plants. The comparative method applied to gene function and genetic controls that determine morphology will vastly increase the power of evo–devo to explain both evolutionary mechanisms and the history of evolutionary change.

Relatively inexpensive 454 Life Sciences[45] and Solexa sequencing can detect virtually all mRNAs in a tissue, so expressed genes are known, and microarrays can measure their relative abundances. At present, in non-model organisms, gene function is often assumed to resemble that of closely related genes in model organisms, but improved reverse-genetics methods, such as VIGS (virus-induced gene silencing)[46] and TILLING (targeting induced local lesions in genomes), can downregulate genes to demonstrate function directly[47].

Phylogenetic footprinting between species identifies conserved non-coding DNA segments that probably have shared protein-binding sites that are important for regulating gene expression[48]. Segments shared by distant species with similar, homologous morphologies versus segments shared among taxa with differing morphologies should reveal similarities and differences in transcription-factor-binding sites, explaining inputs to gene expres sion that result in various morphologies. Surface plasmon resonance and other methods can measure equilibrium and kinetics constants for protein–protein[49] and protein–DNA interactions[50] on short DNA segments and perhaps on promoters of a few thousand base pairs in length; this potentially allows measurement of regulatory outputs of the proteome and inputs to gene expression. In combination, these methods should greatly facilitate elucidation and comparison of genetic control networks in non-model organisms, vastly increasing the power of evo–devo; however, before these studies are available, phylogenetic analyses of individual gene families and expressed sequence tag/microarray studies of whole flowers and floral organs will continue to provide the most useful data, such as those of the Floral Genome Project[19].

The appearance in the past decade of theories of flower origin, stimulated by developmental genetic data from modern plants, marks a major shift in attempts to solve Darwin's "abominable mystery". By building a model of the common aspects of floral developmental controls and comparing these with common elements of gymnosperm systems, we can build a picture of the genetic architecture underpinning floral structure in primitive angiosperms and test theories of how floral systems could have arisen[19]. This could lead to the realization that the fossils we need for understanding angiosperm origins may already be known. Incremental fossil discoveries should allow increasingly complete reconstructions of currently poorly known extinct taxa, which may then be included in phylogenetic analyses, but a palaeobotanical *deus ex machina* is possible at any time if a fossil is discovered that illustrates intermediate steps in the evolution of critical angiosperm attributes, such as the carpel with its included ovules or the angiosperm stamen with its specialized structure.

1. Frohlich, M. W. in *Developmental Genetics of the Flower* (eds Soltis, D. E., Leebens-Mack, J. H. & Soltis, P. S.) 63–127 (Academic, San Diego, CA, 2006).
2. Crane, P. R., Friis, E. M. & Pedersen, K. R. The origin and early diversification of angiosperms. *Nature* **374**, 27–33 (1995).
3. Soltis, P. S., Endress, P. K., Chase, M. W. & Soltis, D. E. *Angiosperm Phylogeny and Evolution* (Sinauer, Sunderland, MA, 2005).
4. Saarela, J. M. *et al.* Hydatellaceae identified as a new branch near the base of the angiosperm phylogenetic tree. *Nature* **446**, 312–315 (2007).
5. Rudall, P. J. *et al.* Morphology of Hydatellaceae, an anomalous aquatic family recently recognized as an early-divergent angiosperm lineage. *Am. J. Bot.* **94**, 1073–1092 (2007).
6. Friis, E. M., Pedersen, K. R. & Crane, P. R. Cretaceous angiosperm flowers: Innovation and evolution in plant reproduction. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **232**, 251–293 (2006).
7. Rydin, C., Pedersen, K. R., Crane, P. R. & Friis, E.-M. Former diversity of *Ephedra* (Gnetales): Evidence from Early Cretaceous seeds from Portugal and North America. *Ann. Bot. (Lond.)* **98**, 123–140 (2006).
8. Friis, E. M., Doyle, J. A., Endress, P. K. & Leng, Q. *Archaefructus*—angiosperm precursor or specialized early angiosperm? *Trends Plant Sci.* **8**, 369–373 (2003).
9. Ji, Q. *et al.* Early Cretaceous *Archaefructus eoflora* sp. nov. with bisexual flowers from Beipiao, Western Liaoning, China. *Acta Geol. Sin. Engl. Edn* **78**, 883–896 (2004).

10. Anderson, J. M. & Anderson, H. M. *Heyday of the Gymnosperms: Systematics and Biodiversity of the late Trassic Molteno Fructifications* (National Botanical Institute, Pretoria, 2003).
11. Crane, P. R. Phylogenetic analysis of seed plants and the origin of angiosperms. *Ann. Mo. Bot. Gard.* **72,** 716–793 (1985).
12. Doyle, J. A. & Donoghue, M. J. Seed plant phylogeny and the origin of angiosperms: an experimental cladistic approach. *Bot. Rev.* **52,** 321–431 (1986).
13. Doyle, J. A. Seed plant phylogeny and the relationships of Gnetales. *Int. J. Plant Sci.* **157,** S3–S39 (1996).
14. Burleigh, J. G. & Mathews, S. Assessing among-locus variation in the inference of seed plant phylogeny. *Int. J. Plant Sci.* **168,** 111–124 (2007).
15. Hilton, J. & Bateman, R. M. Pteridosperms are the backbone of seed-plant phylogeny. *J. Torrey Bot. Soc.* **133,** 119–168 (2006).
16. Bateman, R. M., Hilton, J. & Rudall, P. J. Morphological and molecular phylogenetic context of the angiosperms: contrasting the 'top-down' and 'bottom-up' approaches used to infer the likely characteristics of the first flowers. *J. Exp. Bot.* **13,** 3471–3503 (2006).
17. Doyle, J. A. Seed ferns and the origin of angiosperms. *J. Torrey Bot. Soc.* **133,** 169–209 (2006).
18. Doyle, J. A. & Endress, P. K. Morphological phylogenetic analysis of basal angiosperms: Comparison and combination with molecular data. *Int. J. Plant Sci.* **161,** S121–S153 (2000).
19. Soltis, D. E. *et al.* The evolving floral genome: A history of genome-wide duplications and shifting patterns of gene expression. *Trends Plant Sci.* **12,** 358–367 (2007).
20. Taylor, D. W. *et al.* Biogeochemical evidence for the presence of the angiosperm molecular fossil oleanane in Paleozoic and Mesozoic non-angiospermous fossils. *Paleobiology* **32,** 179–190 (2006).
21. Auras, S. *et al.* Aromatized arborane/fernane hydrocarbons as biomarkers for *Cordaites. Naturwissenschaften* **93,** 616–621 (2006).
22. Stebbins, G. L. *Flowering Plants Evolution above the Species Level* (Harvard Univ. Press, Cambridge, MA, 1974).
23. Retallack, G. & Dilcher, D. L. Arguments for a glossopterid ancestry of angiosperms. *Paleobiology* **71,** 54–67 (1981).
24. Friedman, W. E. & Williams, J. H. Modularity of the angiosperm female gametophyte and its bearing on the early evolution of endosperm in flowering plants. *Evolution Int. J. Org. Evolution* **57,** 216–230 (2003).
25. Hughes, N. F. 1994. *The Enigma of Angiosperm Origins* (Cambridge Univ. Press, Cambridge, 1994).
26. Krassilov, V. A. *Angiosperm Origins: Morphological and Ecological Aspects* (Pensoft Publishers, Sofia, Bulgaria, 1997).
27. Litt, A. Evaluation of A-function: evidence from the *APETALA1* and *APETALA2* gene lineages. *Int. J. Plant Sci.* **168,** 73–91 (2007).
28. Albert, V. A., Gustafsson, M. H. G. & Di Laurenzio, L. in *Molecular Systematics of Plants II: DNA Sequencing* (eds Soltis, D. E., Soltis, P. S. & Doyle, J. J.) 349–374 (Kluwer Academic, Dordrecht, 1998).
29. Kramer, E. M., Jaramillo, M. A. & Di Stilio, V. S. Patterns of gene duplication and functional evolution during the diversification of the *AGAMOUS* subfamily of MADS box genes in angiosperms. *Genetics* **166,** 1011–1023 (2004).
30. Zahn, L. M. *et al.* Conservation and divergence in the *AGAMOUS* subfamily of MADS-box genes: evidence of independent sub- and neofunctionalization events. *Evol. Dev.* **8,** 30–45 (2006).
31. Skinner, D. J., Hill, T. A. & Gasser, C. S. Regulation of ovule development. *Plant Cell* **16,** S32–S45 (2004).
32. Buzgo, M., Soltis, P. S. & Soltis, D. E. Floral developmental morphology of *Amborella trichopoda* (Amborellaceae). *Int. J. Plant Sci.* **165,** 925–947 (2004).
33. Theissen, G. *et al.* in *Developmental Genetics and Plant Evolution* (eds Cronk, Q. C. B., Bateman, R. M. & Hawkins, J. A.) 173–206 (Taylor & Francis, London, 2002).
34. Theissen, G. & Becker, A. Gymnosperm orthologues of class B floral homeotic genes and their impact on understanding flower origin. *Crit. Rev. Plant Sci.* **23,** 129–148 (2004).
35. Baum, D. A. & Hileman, L. C. in *Flowering and its Manipulation* (ed. Ainsworth, C.) 3–27 (Blackwell, Oxford, 2006).
36. Frohlich, M. W. & Parker, D. S. The Mostly Male theory of flower evolutionary origins: from genes to fossils. *Syst. Bot.* **25,** 155–170 (2000).
37. Frohlich, M. W. in *Developmental Genetics and Plant Evolution* (eds Cronk, Q. C. B. & Bateman, R. M. & Hawkins, J. A.) 85–108 (Taylor & Francis, London, 2002).
38. Frohlich, M. W. An evolutionary scenario for the origin of flowers. *Nature Rev. Genet.* **4,** 559–566 (2003).
39. Vázquez-Lobo, A. *et al.* Characterization of the expression patterns of LEAFY/FLORICAULA and NEEDLY orthologs in female and male cones of the conifer genera *Picea*, *Podocarpus* and *Taxus*: implications for current evo-devo hypotheses for gymnosperms. *Evol. Dev.* **9,** 446–459 (2007).
40. Meyen, S. V. Origin of the angiosperm gynoecium by gametoheterotopy. *Bot. J. Linn. Soc.* **97,** 171–178 (1988).
41. Kim, S. *et al.* Sequence and expression studies of A-, B-, and C-class Mads-box homologues in *Eupomatia* (Eupomatiaceae): support for the bracteate origin of the calyptra. *Int. J. Plant Sci.* **166,** 185–198 (2005).
42. Jaramillo, M. A. & Kramer, E. M. The role of developmental genetics in understanding homology and morphological evolution in plants. *Int. J. Plant Sci.* **168,** 61–72 (2007).
43. Long, J. & Barton, M. K. Initiation of axillary and floral meristems in *Arabidopsis. Dev. Biol.* **218,** 341–353 (2000).
44. Causier, B. *et al.* Evolution in action: Following function in duplicated floral homeotic genes. *Curr. Biol.* **15,** 1508–1512 (2005).
45. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437,** 376–380 (2005).
46. Burch-Smith, T. M., Anderson, J. C., Martin, G. B. & Dinesh-Kumar, S. P. Applications and advantages of virus-induced gene silencing for gene function studies in plants. *Plant J.* **39,** 734–746 (2004).
47. David-Schwartz, R. & Sinha, N. Evolution and development in plants: bridging the gap. *Int. J. Plant Sci.* **168,** 49–59 (2007).
48. De Bodt, S., Theissen, G. & Van de Peer, Y. Promoter analysis of MADS-box genes in eudicots through phylogenetic footprinting. *Mol. Biol. Evol.* **23,** 1293–1303 (2006).
49. Boozer, C. *et al.* Looking towards label-free biomolecular interaction analysis in a high-throughput format: a review of new surface plasmon resonance technologies. *Curr. Opin. Biotechnol.* **17,** 400–405 (2006).
50. Nalefski, E. A., Nebelitsky, E., Lloyd, J. A. & Gullans, S. R. Single-molecule detection of transcription factor binding to DNA in real time: Specificity, equilibrium, and kinetic parameters. *Biochemistry* **45,** 13794–13806 (2006).
51. Pryer, K. M. *et al.* Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature* **409,** 618–622 (2001).
52. Klavins, S. D., Taylor, T. N. & Taylor, E. L. Anatomy of *Umkomasia* (Corystospermales) from the Triassic of Antarctica. *Am. J. Bot.* **89,** 664–676 (2002).
53. Moore, M. J. *et al.* Using plastid-genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proc. Natl. Acad. Sci. USA* **104,** 19363–19368 (2007).

Acknowledgements M.W.F. thanks the National Science Foundation (USA) for supporting work in this area. We thank J. A. Doyle, E.-M. Friis, P. S. Soltis, R. M. Bateman, P. Kenrick, D. E. Soltis and J. Hilton for commenting on the manuscript, and J. Trager and Huntington Gardens for *Welwitschia* materials.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence should be addressed to M.W.F. (m.frohlich@kew.org).

Vol 450|20/27 December 2007|doi:10.1038/nature06343

ARTICLES

Whales originated from aquatic artiodactyls in the Eocene epoch of India

J. G. M. Thewissen[1], Lisa Noelle Cooper[1,2], Mark T. Clementz[3], Sunil Bajpai[4] & B. N. Tiwari[5]

Although the first ten million years of whale evolution are documented by a remarkable series of fossil skeletons, the link to the ancestor of cetaceans has been missing. It was known that whales are related to even-toed ungulates (artiodactyls), but until now no artiodactyls were morphologically close to early whales. Here we show that the Eocene south Asian raoellid artiodactyls are the sister group to whales. The raoellid *Indohyus* is similar to whales, and unlike other artiodactyls, in the structure of its ears and premolars, in the density of its limb bones and in the stable-oxygen-isotope composition of its teeth. We also show that a major dietary change occurred during the transition from artiodactyls to whales and that raoellids were aquatic waders. This indicates that aquatic life in this lineage occurred before the origin of the order Cetacea.

Phylogenetic analyses of molecular data on extant animals strongly support the notion that hippopotamids are the closest relatives of cetaceans (whales, dolphins and porpoises)[1–3]. In spite of this, it is unlikely that the two groups are closely related when extant and extinct artiodactyls are analysed, for the simple reason that cetaceans originated about 50 million years (Myr) ago in south Asia, whereas the family Hippopotamidae is only 15 Myr old, and the first hippopotamids to be recorded in Asia are only 6 Myr old[4]. However, analyses of fossil clades have not resolved the issue of cetacean relations. Proposed sister groups ranged from the entire artiodactyl order[5,6], to the extinct early ungulates mesonychians[7], to an anthracotheroid clade[8] (which included hippopotamids), to weakly supporting hippopotamids (to the exclusion of anthracotheres[9,10]).

The middle Eocene artiodactyl family Raoellidae[11–14] is broadly coeval with the earliest cetaceans, and both are endemic to south Asia. Raoellids, as a composite consisting of several genera, have been added to some phylogenetic analyses[5,10], but no close relation to whales was found because raoellid fossils were essentially limited to dental material[11–14]. We studied new dental, cranial and postcranial material for *Indohyus*, a middle Eocene raoellid artiodactyl from Kashmir, India (Fig. 1). All fossils of *Indohyus* were collected at a middle Eocene bone bed extending for about 50 m at the locality Sindkhatudi in the Kalakot region of Kashmir on the Indian side of the Line of Control. Our analysis identifies raoellids as the sister group to cetaceans and bridges the morphological divide that separated early cetaceans from artiodacyls. This has profound implications for the character transformations near the origin of cetaceans and the cladistic definition of Cetacea, and identifies the habitat in which whales originated. Taken together, our findings lead us to propose a new hypothesis for the origin of whales.

Cetaceans and raoellids are sister groups

To investigate the importance of raoellids in cetacean phylogeny, we excluded raoellids from our initial phylogenetic analysis of artiodactyls plus cetaceans. Our data set differed from previous analyses[10] by the addition of several archaic anthracotheres, and some corrected scores for pakicetid cetaceans. This analysis found stronger support for hippopotamid–cetacean sister-group relations than the previous analysis[10], consistent with molecular studies[1–3]. However, the base of the artiodactyl cladogram is poorly resolved (see Supplementary Information for details on phylogenetic runs). In a second cladistic analysis (Fig. 2), we added the raoellids *Khirtharia* and *Indohyus* as well as several archaic ungulate groups (condylarths) and found that raoellids and cetaceans are sister groups and that they are the basal node in the Cetacea/Artiodactyla clade, consistent with some previous analyses that used different character sets[5,6]. Our analysis is the first to show that raoellids are the sister group to cetaceans, resolving the biogeographic conundrum and closing the temporal gap between cetaceans and their sister. Relations between most artiodactyl families higher in the tree are poorly resolved, and our data lack implications for the relations between these families. Our analysis strongly argues that raoellids and cetaceans are more closely related to each other than either is to hippopotamids.

Indohyus shares with cetaceans several synapomorphies that are not present in other artiodactyls. Most significantly, *Indohyus* has a thickened medial lip of its auditory bulla, the involucrum (Figs 1 and 3), a feature previously thought to be present exclusively in cetaceans. Involucrum size varies among cetaceans, but the relative thickness of medial and lateral walls of the tympanic of *Indohyus* is clearly within the range of that of cetaceans and is well outside the range of other cetartiodactyls (Fig. 3). Other significant derived similarities between *Indohyus* and cetaceans include the anteroposterior arrangement of incisors in the jaw, and the high crowns in the posterior premolars.

Characterizing Cetacea

Until now, the involucrum was the only character occurring in all fossil and recent cetaceans but in no other mammals[5,15,16]. Identification of the involucrum in *Indohyus* calls into question what it is to be a cetacean: it requires either that the concept of Cetacea be expanded to include *Indohyus* or that the involucrum cease to characterize cetaceans. We argue that the content of Cetacea should remain stable and include Pakicetidae, Ambulocetidae, Remingtonocetidae, Protocetidae, Basilosauridae, Mysticeti and Odontoceti[6,7,9,10,17]. Thus, Cetacea remains a monophyletic group, whereas Artiodactyla remains a paraphyletic group (because Raoellidae are included but Cetacea are excluded). An alternative classification would render both Cetacea and Artiodactyla monophyletic by including Raoellidae in Cetacea and

[1]Department of Anatomy, Northeastern Ohio Universities College of Medicine, Rootstown, Ohio 44272, USA. [2]School of Biomedical Sciences, Kent State University, Kent, Ohio 44242, USA. [3]Department of Geology and Geophysics, University of Wyoming, Laramie, Wyoming 82071, USA. [4]Department of Earth Sciences, Indian Institute of Technology, Roorkee, Uttarakhand 247 667, India. [5]Wadia Institute of Himalayan Geology, Dehra Dun, Uttarakhand 248 001, India.

by limiting Artiodactyla to those clades one node above the raoellid/cetacean node (Fig. 2). We do not prefer this classification because it causes instability by significantly altering the traditional content of both Artiodactyla and Cetacea.

Characters identified as synapomorphies for Cetacea in some of our most parsimonious trees include: long external auditory meatus, double-rooted P3/, lack of P4/ protocone, M1-2/ metacones present but small, and lack of M1-2/ hypocone. None of these features characterize all modern and extinct cetaceans; the dental characters, for instance, cannot be scored in toothless mysticetes. In addition, all of these characters are found in some mammals unrelated to cetaceans.

We attach particular importance to two character complexes that characterize basal cetaceans, constitute synapomorphic suites for the order, and are of great functional importance. All fossil and recent cetaceans differ from most other mammals in the reduction of crushing basins on their teeth: there are no trigonid and talonid basins in the lower molars, and the trigon basin of the upper molars is very small (for example in pakicetids and ambulocetids) or absent. Crushing basins are large in raoellids (Fig. 1a, b) and other basal ungulates. This implies that a major change in dental function occurred at the origin of cetaceans, probably related to dietary change at the origin[18]. Reduced crushing basins also occur in mesonychids, archaic ungulates long thought to be closely related to cetaceans. However, mesonychian molars have wear facets very unlike those of cetaceans[7,18], whereas wear facets in raoellids are more similar to wear facets in early cetaceans[14].

The second character complex that identifies cetaceans is the shape of the postorbital and temporal region of the skull. In early cetaceans, this region is long and narrow[19]. This affects the sense organs: the olfactory peduncle is long and narrow and the orbits are set close together near the roof of the skull. It also affects oral function, the nasopharyngeal duct is narrow, and the out-lever of the masticatory muscles is long, increasing the closing speed of the jaws. We speculate that the importance of different sense organs was related to these changes, or that changing diet led to a change in food-processing organs.

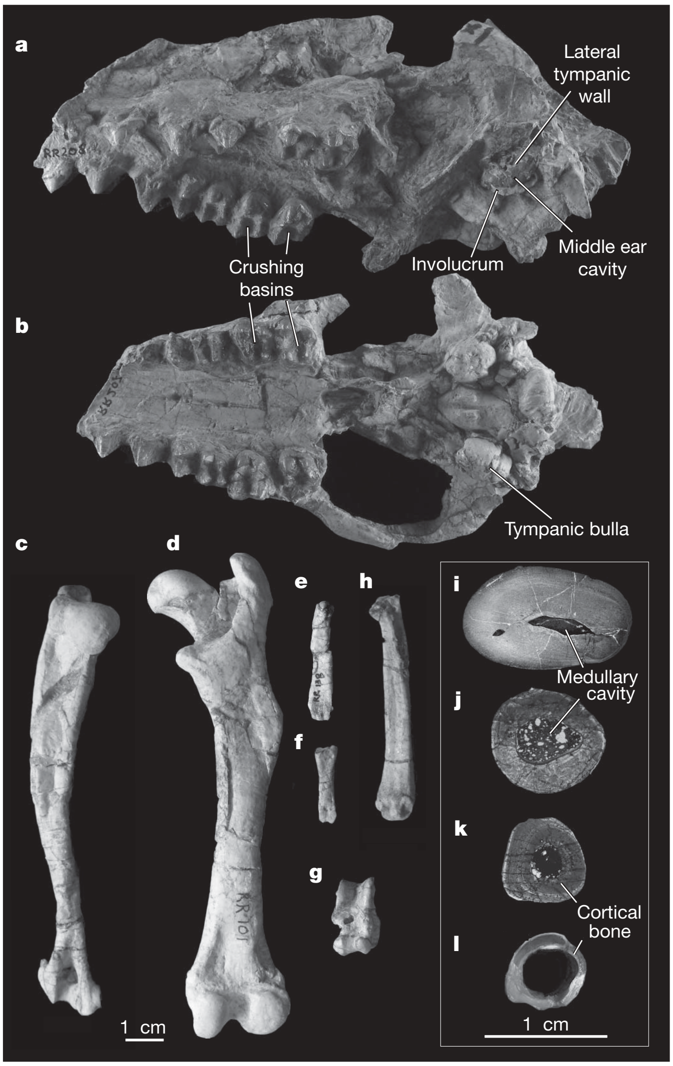


Figure 1 | Osteology of *Indohyus* and cross-sections of long bones of Eocene cetartiodactyls. **a**, **b**, Oblique lateral view of skull RR 208 (**a**) and ventral view of skull RR 207 (**b**). **c–h**, Posterior views of humerus (RR 149, **c**) and femur (RR 101, **d**), plantar views of metacarpal (RR 138, **e**) and proximal manual phalanx (RR 19, **f**), dorsal view of astragalus (RR 224, **g**), and posterior view of metatarsal (RR 139, **h**). **i–l**, Histological mid-shaft sections for humerus of the pakicetid *Ichthyolestes* (H-GSP 96227, **i**), humerus of *Indohyus* (RR 157, **j**), femur of *Indohyus* (RR 42, **k**) and femur of the artiodactyl *Cainotherium* (IVAU unnum, **l**). Both scale bars are 1 cm; the scale bar near **d** goes with **a–h**, and that near **l** goes with **i–l**.

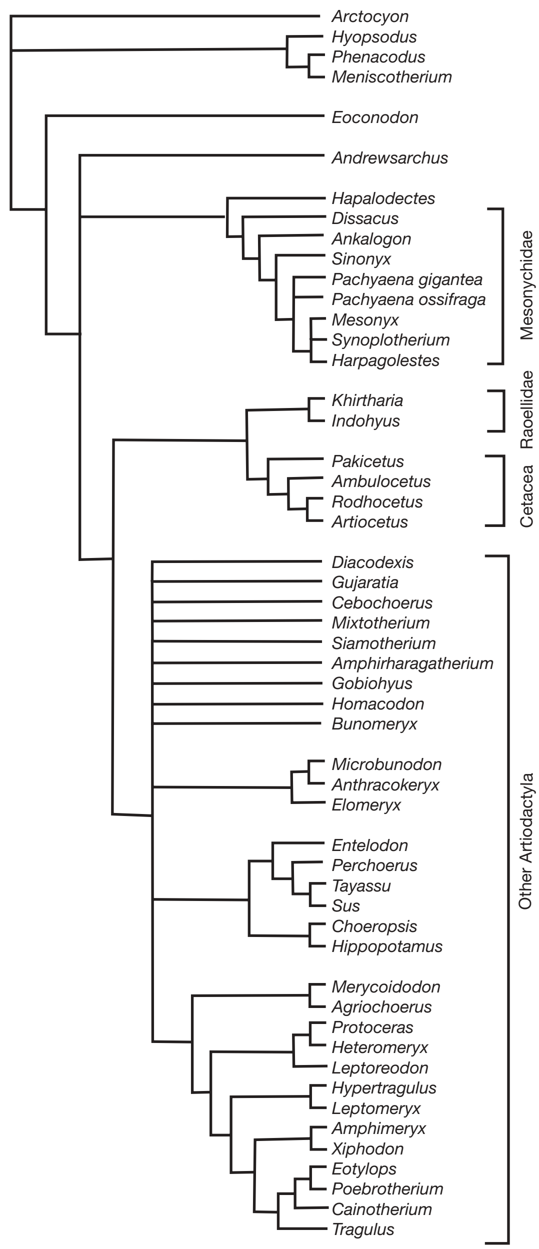


Figure 2 | Phylogeny of artiodactyls, cetaceans and archaic ungulates. The figure shows a consensus cladogram produced by heuristic searches with PAUP (random addition sequence, 1,000 repetitions), using a published data set[10]. See Supplementary Information for further details.

Indohyus was aquatic

Behaviourally, the earliest whales (pakicetids) were aquatic waders[5,20–23]. This led us to investigate whether *Indohyus* was aquatic too. Cortical bone thickness in secondarily aquatic tetrapods is commonly increased at the expense of the medullary cavity, a pattern called osteosclerosis[24]. Osteosclerosis occurs in early whales[20,21,25], manatees[26,27], sea otters[28], *Hippopotamus*[29], beavers[29], pinnipeds[29] and Mesozoic marine reptiles[26,30]. Osteosclerosis provides ballast that allows some aquatic taxa to be bottom walkers (hippopotamids) and others to maintain neutral buoyancy in water (manatees)[30]. Histological sections indicate that the limb bones of *Indohyus* are also osteosclerotic (Fig. 1i–l), in a similar manner to those of pakicetid cetaceans. Our survey of cortical bone thickness in the limb bones of terrestrial artiodactyls shows that this pattern is unusual for that order: in mid-bone cross-sections of the femur, the medullary cavity makes up between 0.60 and 0.75 of the width of the bone, whereas in aquatic mammals the values are lower (*Hippopotamus*, 0.55; pakicetids and ambulocetids, 0.25–0.57). In *Indohyus* this ratio is 0.42, suggesting that *Indohyus* was osteosclerotic and thus aquatic. We interpret the limb osteosclerosis of *Indohyus* to be related to bottom walking and not to slow swimming, because the limbs are gracile and not modified into paddles.

To investigate further the hypothesis that *Indohyus* was aquatic, we studied the stable isotopes of its enamel, a tissue relatively resistant to preburial and postburial alteration of isotopic composition[31]. Enamel δ^{18}O values are influenced by the oxygen isotope composition of the food and water ingested by an animal as well as by certain physiological processes (such as sweating, panting and respiration)[32]. For aquatic species, the flux of environmental water by means of direct ingestion and transcutaneous exchange overwhelms all other oxygen sources[33] and can cause the enamel δ^{18}O values of freshwater taxa (for example *Hippopotamus*) to be 2–3‰ lower than those for terrestrial mammals[33,34]. Mean δ^{18}O values for four individuals of *Indohyus* are at least 2‰ lower than those for our comparative sample of Eocene terrestrial and semi-aquatic mammals from formations of India and Pakistan of similar or slightly older age (Fig. 4)[22,23]. Although not representative of the specific deposits from which *Indohyus* was collected, oxygen isotope values for each ecological type from these sites (namely terrestrial, 24–28‰; semi-aquatic, 23‰ or less) are surprisingly consistent regardless of age or location (Fig. 4). This suggests that temporal and spatial variation in environmental isotope values was relatively minor and was most probably insufficient to account for the extremely low δ^{18}O we have reported for *Indohyus*. We did not recover tooth material of other mammals at the *Indohyus* site; until such material can be analysed, the most consistent interpretation is that these low values are a result of the aquatic habits of this species.

Supporting evidence for a semi-aquatic life of *Indohyus* comes from examination of its inter-individual variation in δ^{18}O values. The overwhelming influx and mitigating influence of isotopically homogeneous environmental water causes the variation in individual δ^{18}O values for semi-aquatic and aquatic species (s.d. < 0.5‰) to be much lower than that of terrestrial species (s.d. > 1.0‰)[33]. This line of evidence is especially relevant in our study because it does not require an approximation of the mean environmental isotope values for a site through analysis of the associated fauna. Variation in δ^{18}O values for *Indohyus* (s.d. $= 0.4$‰) is extremely low and when compared with species of sufficient sample size ($n > 3$) it is similar to that of the semi-aquatic archaeocete *Pakicetus*. Given that the influence of physiological and environmental factors on body water δ^{18}O values is more strongly felt at smaller body sizes[35], this low level of variation is particularly compelling for *Indohyus*, with a body mass of less than 50 kg.

To explore the diet of *Indohyus* we studied carbon isotopes. Enamel δ^{13}C values are defined by the carbon isotope composition of an animal's diet and can be used to identify the food webs and resources used by an animal[36]. The δ^{13}C values of primary producers at the base of aquatic and terrestrial food webs overlap, but values for freshwater phytoplankton are typically depleted in ^{13}C relative to freshwater macrophytes[37], and both types of aquatic producer are depleted in ^{13}C relative to terrestrial C_3 plants[37–39]. Consumers foraging within food webs fuelled by freshwater phytoplankton (for example freshwater and brackish-water foraging Eocene whales) typically have lower δ^{13}C values than species foraging on aquatic macrophytes[37] or on terrestrial resources[33] (Fig. 4). Enamel δ^{13}C values for *Indohyus* are higher than those for most early cetaceans and are most similar to the δ^{13}C values in enamel for terrestrial mammals from early and middle Eocene deposits in India and Pakistan. *Indohyus* could have been feeding on land or in water, but it was clearly eating something different from archaeocetes such as *Pakicetus* and *Ambulocetus*. If the large crushing basins in the molars of *Indohyus* were used for processing vegetation, these δ^{13}C values in enamel could come from the ingestion of terrestrial plants or aquatic macrophytes. Alternatively, a more ominivorous diet would suggest that *Indohyus* might have foraged on benthic, aquatic invertebrates in

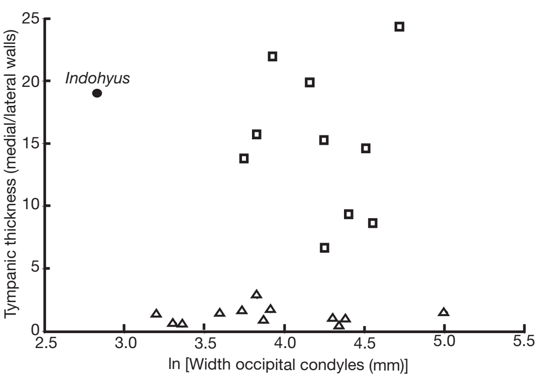


Figure 3 | Plot of the ratio of the thickness of the medial tympanic wall to that of the lateral tympanic wall against the natural logarithm of the width across occipital condyles, showing that the ratio in *Indohyus* is similar to that in cetaceans. In cetaceans (open squares), the medial tympanic wall is inflated and called the involucrum, and the lateral tympanic wall is thinned and called the tympanic plate. In artiodactyls (open triangles), the medial and lateral tympanic walls are more similar in thickness, causing values on the *y* axis to be closer to 1. See Supplementary Information for further details.

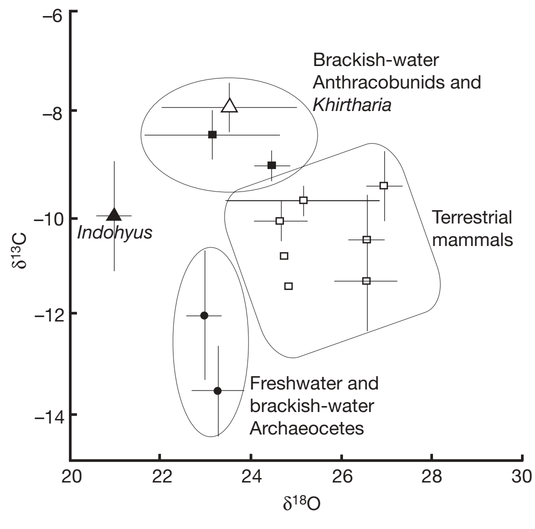


Figure 4 | Bivariate plot of δ^{18}O and δ^{13}C values for enamel samples of early and middle Eocene mammals from India and Pakistan. Results are shown as means ± s.d. for the sample population. Filled triangle, *Indohyus*; open triangle, *Khirtharia*; open squares, terrestrial mammals; filled squares, brackish-water anthracobunids; filled circles, freshwater/brackish-water archaeocetes. See Supplementary Information for details.

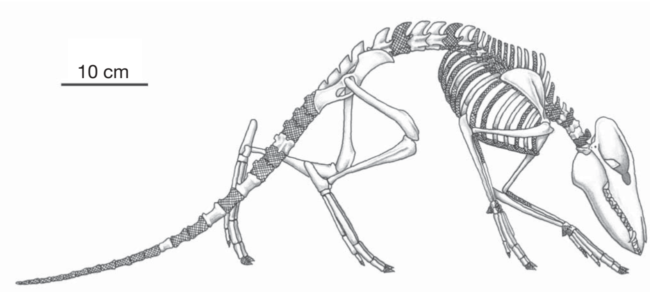


Figure 5 | Skeletal reconstruction of *Indohyus*. Hatched elements are reconstructed on the basis of related taxa.

freshwater systems. Although we cannot exclude the possibility of aquatic foraging by *Indohyus*, $\delta^{13}C$ values in enamel do suggest that the diet of *Indohyus* differed significantly from that of Eocene whales. A more refined interpretation of the dietary preferences of *Indohyus* will require a study of tooth wear and tooth morphology.

Evolutionary hypothesis for whale origins

Indohyus was a small, stocky artiodactyl, roughly the size of the raccoon *Procyon lotor* (Fig. 5). It was not an adept swimmer; instead it waded in shallow water, with its heavy bones providing ballast to keep its feet anchored. *Indohyus* may have fed on land, although a specialized aquatic diet is also possible.

The modern artiodactyl morphologically most similar to *Indohyus* is probably the African mousedeer *Hyemoschus aquaticus*. *Hyemoschus* lives near streams and feeds on land, but flees into the water when danger occurs[40]. *Indohyus* had more pronounced aquatic specializations than *Hyemoschus* does, and it probably spent a considerably greater amount of time in the water either for protection or when feeding. As indicated by the evidence from stable isotopes, *Indohyus* spent most of its time in the water and either came onshore to feed on vegetation (as the modern *Hippopotamus* does) or foraged on invertebrates or aquatic vegetation in the same way that the modern muskrat *Ondatra* does.

Raoellids are the sister group to cetaceans, and this implies that aquatic habitats originated before the Order Cetacea. The great evolutionary change that occurred at the origin of cetaceans is thus not the adoption of an aquatic lifestyle. Here we propose that dietary change was the event that defined cetacean origins; this is consistent with the cranial and dental synapomorphies identified. Molars of *Indohyus* are markedly different from those of pakicetids, and it is widely assumed that pakicetids ate aquatic prey[18,23].

Our working hypothesis for the origin of whales is that raoellid ancestors, although herbivores or omnivores on land, took to fresh water in times of danger. Aquatic habits were increased in *Indohyus* (as suggested by osteosclerosis and oxygen isotopes), although it did not necessarily have an aquatic diet (as suggested by carbon isotopes). Cetaceans originated from an *Indohyus*-like ancestor and switched to a diet of aquatic prey. Significant changes in the morphology of the teeth, the oral skeleton and the sense organs made cetaceans different from their ancestors and unique among mammals.

METHODS SUMMARY

We chose an existing character matrix[10] as the basis for our phylogenetic analysis. We corrected scores of some of the taxa, and made some changes in the taxa included. Details on these taxa, the rationale for using them, and their scores are given in Methods and in Supplementary Table 1.

Tympanic wall thickness was investigated to address the presence of the involucrum quantitatively. We calculated the ratio of medial tympanic wall thickness divided by lateral tympanic wall thickness (see Methods and Supplementary Table 2). Bone histology was studied to investigate the presence of osteosclerosis. It was quantified as the ratio of medullary cavity width divided by bone width in the mediolateral plane (see Methods and Supplementary Table 3).

For analysis of stable isotopes, we prepared powders by following published methods (see Methods and Supplementary Tables 4 and 5). Multiple samples were collected for each species to provide an estimate of population means for carbon and oxygen isotope values[33]. About 5 mg of enamel powder was collected from each specimen (tooth) for study of carbon and oxygen isotope values.

Most fossils of *Indohyus* were collected by the late Indian geologist A. Ranga Rao, who discovered the locality about 25 years ago (acronym RR); additional fossils were collected by S.B. and B.N.T. at the same locality (acronym IITR-SB-Kal-S).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 June; accepted 3 October 2007.

1. Milinkovitch, M. C., Bérubé, M. & Palsbøl, P. J. in *The Emergence of Whales: Evolutionary Patterns in the Origin of Cetacea* (ed. Thewissen, J. G. M.) 113–131 (Plenum, New York, 1998).
2. Nikaido, M., Rooney, A. P. & Okada, N. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: Hippopotamuses are the closest extant relatives of whales. *Proc. Natl Acad. Sci. USA* **96,** 10261–10266 (1999).
3. Gatesy, J. & O'Leary, M. A. Deciphering whale origins with molecules and fossils. *Trends Ecol. Evol.* **16,** 562–570 (2001).
4. Boisserie, J.-R., Lihoreau, F. & Brunet, M. Origins of Hippopotamidae (Mammalia, Cetartiodactyla): towards resolution. *Zool. Scr.* **34,** 119–143 (2005).
5. Thewissen, J. G. M., Williams, E. M., Roe, L. J. & Hussain, S. T. Skeletons of terrestrial cetaceans and the relationships of whales to artiodactyls. *Nature* **413,** 277–281 (2001).
6. Theodor, J. M. & Foss, S. E. Deciduous dentitions of Eocene cebochoerid artiodactyls and cetartiodactyl relationships. *J. Mammal. Evol.* **12,** 161–181 (2005).
7. O'Leary, M. A. in *The Emergence of Whales: Evolutionary Patterns in the Origin of Cetacea* (ed. Thewissen, J. G. M.) 133–161 (Plenum, New York, 1998).
8. Boisserie, J.-R., Lihoreau, F. & Brunet, M. The position of Hippopotamidae within Cetartiodactyla. *Proc. Natl Acad. Sci. USA* **102,** 1537–1541 (2005).
9. Geisler, J. H. & Uhen, M. D. Morphological support for a close relationship between hippos and whales. *J. Vertebr. Paleontol.* **23,** 991–996 (2003).
10. Geisler, J. H. & Uhen, M. D. Phylogenetic relationships of extinct Cetartiodactyls: results of simultaneous analyses of molecular, morphological, and stratigraphic data. *J. Mammal. Evol.* **12,** 145–160 (2005).
11. Ranga Rao, A. New mammals from Murree (Kalakot zone) of the foot-hills near Kalakot, J & K State. *J. Geol. Soc. India* **12,** 125–134 (1971).
12. Ranga Rao, A. & Misra, V. N. On the new Eocene mammal localities in the Himalayan foot-hills. *Himalayan Geol.* **11,** 422–428 (1981).
13. Kumar, K. & Sahni, A. Eocene mammals from the Upper Subathu Group, Kashmir Himalaya, India. *J. Vertebr. Paleontol.* **5,** 153–168 (1985).
14. Thewissen, J. G. M., Williams, E. M. & Hussain, S. T. Eocene mammal faunas from northern Indo-Pakistan. *J. Vertebr. Paleontol.* **21,** 347–366 (2001).
15. Luo, Z. in *The Emergence of Whales: Evolutionary Patterns in the Origin of Cetacea* (ed. Thewissen, J. G. M.) 269–301 (Plenum, New York, 1998).
16. Luo, Z. & Gingerich, P. D. Terrestrial Mesonychia to aquatic Cetacea: transformation of the basicranium and evolution of hearing in whales. *Univ. Mich. Pap. Paleontol.* **31,** 1–98 (1999).
17. Thewissen, J. G. M & Williams, E. M. The early radiations of Cetacea (Mammalia): evolutionary pattern and developmental correlations. *Annu. Rev. Ecol. Syst.* **33,** 73–90 (2002).
18. O'Leary, M. A. & Uhen, M. D. The time of origin of whales and the role of behavioral changes in the terrestrial–aquatic transition. *Paleobiology* **25,** 534–556 (1999).
19. Nummela, S., Hussain, S. T. & Thewissen, J. G. M. Cranial anatomy of Pakicetidae (Cetacea, Mammalia). *J. Vertebr. Paleontol.* **26,** 746–759 (2006).
20. Gray, N. M., Kainec, K., Madar, S., Tomko, L. & Wolfe, S. Sink or swim? Bone density as a mechanism for buoyancy control in early cetaceans. *Anat. Rec.: Adv. Int. Anat. Evol. Biol.* **290,** 638–653 (2007).
21. Madar, S. I. The postcranial skeleton of early Eocene pakicetid cetaceans. *J. Paleontol.* **81,** 176–200 (2007).
22. Roe, L. J. *et al.* in *The Emergence of Whales: Evolutionary Patterns in the Origin of Cetacea* (ed. Thewissen, J. G. M.) 399–422 (Plenum, New York, 1998).
23. Clementz, M. T., Goswami, A., Gingerich, P. D. & Koch, P. L. Isotopic records from early whales and sea cows: contrasting patterns of ecological transition. *J. Vertebr. Paleontol.* **26,** 355–370 (2006).
24. Francillon-Vieillot, H. *et al.* in *Skeletal Biomineralization Patterns, Processes, and Evolutionary Trends* (ed. Carter, J. G.) 471–530 (Van Nostrand Reinhold, New York, 1990).
25. de Buffrénil, V., Ricqlès, A., Ray, C. E. & Domning, D. P. Bone histology of the ribs of the archaeocetes (Mammalia: Cetacea). *J. Vertebr. Paleontol.* **10,** 455–466 (1990).
26. Kaiser, H. E. Untersuchungen zur vergleichenden Osteologie der Fossilen und rezenten Pachyostosen. *Palaeontographica A* **114,** 113–196 (1960).
27. Domning, D. P. & de Buffrénil, V. Hydrostasis in the Sirenia: quantitative data and functional interpretations. *Mar. Mamm. Sci.* **7,** 331–368 (1991).

28. Fish, F. E. & Stein, B. R. Functional correlates of differences in bone density among terrestrial and aquatic genera in the family Mustelidae (Mammalia). *Zoomorphology* **110,** 339–345 (1991).
29. Wall, W. P. The correlation between high limb-bone density and aquatic habitats in recent mammals. *J. Paleontol.* **57,** 197–207 (1983).
30. Taylor, M. A. Functional significance of bone ballast in the evolution of buoyancy control strategies by aquatic tetrapods. *Histor. Biol* **14,** 15–31 (2000).
31. Koch, P. L., Tuross, N. & Fogel, M. L. The effects of sample treatment and diagenesis on the isotopic integrity of carbonate in biogenic hydroxyapatite. *J. Archaeol. Sci.* **24,** 417–429 (1997).
32. Kohn, M. J. Predicting animal δ^{18}O: accounting for diet and physiological adaptation. *Geochim. Cosmochim. Acta* **60,** 4811–4829 (1996).
33. Clementz, M. T. & Koch, P. L. Differentiating aquatic mammal habitat and foraging ecology with stable isotopes in tooth enamel. *Oecologia* **129,** 461–472 (2001).
34. Bocherens, H., Koch, P. L., Mariotti, A., Geraads, D. & Jaeger, J.-J. Isotopic biogeochemistry (^{13}C, ^{18}O) of mammalian enamel from African Pleistocene hominid sites. *Palaios* **11,** 306–318 (1996).
35. Bryant, J. D. & Froelich, P. N. A model of oxygen isotope fractionation in body water of large mammals. *Geochim. Cosmochim. Acta* **59,** 4523–4537 (1995).
36. Jim, S., Ambrose, S. H. & Evershed, R. P. Stable carbon isotopic evidence for differences in the dietary origin of bone cholesterol, collagen and apatite: Implications for their use in palaeodietary reconstruction. *Geochim. Cosmochim. Acta* **68,** 61–72 (2004).
37. Cloern, J. E., Canuel, E. A. & Harris, D. Stable carbon and nitrogen isotope composition of aquatic and terrestrial plants of the San Francisco Bay estuarine system. *Limnol. Oceanogr.* **47,** 713–729 (2002).
38. Osmond, C. B., Valaane, N., Haslam, S. M., Uotila, P. & Roksandic, Z. Comparisons of δ^{13}C values in leaves of aquatic macrophytes from different habitats in Britain and Finland; some implications for photosynthetic processes in aquatic plants. *Oecologia* **50,** 117–124 (1981).
39. O'Leary, M. H. Carbon isotopes and photosynthesis. *Bioscience* **38,** 328–336 (1988).
40. Dubost, G. Un aperçu sur l'écologie du chevrotain africain *Hyemoschus aquaticus* Ogilby, artiodactyle tragulide. *Mammalia* **42,** 1–62 (1978).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the late F. Obergfell for presenting us with the sediment blocks containing *Indohyus* fossils collected by A. Ranga Rao for preparation and study; D. S. N. Raju and N. Raju for facilitating our research; B. Armfield, R. Conley and A. Maas for fossil preparation; J. Dillard for preparing Fig. 5; and J. Geisler and J. Theodor for providing additional information about their cladistic analyses. Laboratory research was funded by the National Science Foundation (NSF) – Earth Sciences (grants to J.G.M.T. and M.T.C.). Collaborative work was funded by the Indian Department of Science and Technology (to S.B.) and the NSF – International Division (to J.G.M.T.) under the Indo-US Scientific Cooperation Program. Laboratory analyses were supported by the Skeletal Biology Research Focus Area of Northeastern Ohio Universities College of Medicine.

Author Contributions J.G.M.T. was responsible for anatomical and systematic study, and scientific synthesis, L.N.C. for systematic and bone density study, M.T.C. for the study of stable isotopes, and S.B. and B.N.T. for geological study and collecting of *Indohyus* and comparative fossil samples.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.G.M.T. (thewisse@neoucom.edu).

METHODS

Systematic study. We chose an existing, published character matrix[10] as the basis for our phylogenetic analysis because this matrix is rich in characters and contains most relevant taxa. From this matrix we deleted those artiodactyls and cetaceans that are geologically young or modern and are well represented by fossil relatives (*Odocoileus*, *Bos*, *Ovis*, *Remingtonocetus*, *Protocetus*, *Georgiacetus*, *Basilosaurus*, *Balaenoptera*, *Physeter*, *Tursiops*, *Delphinapterus*, *Camelus* and *Lama*), as well as perissodactyls and non-ungulate taxa.

To this matrix, we added the anthracotheres *Siamotherium*, *Anthracokeryx* and *Microbunodon*, because they are near the base of the anthracotheroid clade (Anthracotheriidae plus Hippopotamidae) and are sometimes thought to be close to early whales[4]. Scores for these taxa, and the sources on which we based the scores, are listed in Supplementary Table 1 (refs 4, 6, 41–46).

We chose *Gujaratia pakistanensis* (formerly *Diacodexis pakistanensis*)[47] as outgroup for the analyses of cetaceans plus artiodactyls, and we chose *Arctocyon* and *Hyopsodus* as outgroups for the (second) analysis that included all taxa (Fig. 2).

We corrected some of the scores for Pakicetidae[10], because new fossils have been published for this family, in particular cranial material[19] and postcranial material[21]. Corrected scores for pakicetids are also listed in Supplementary Table 1 and were based on original material in the Howard-Geological Survey of Pakistan (H-GSP) collections, currently curated by J.G.M.T.

Raoellidae have been included in several previous phylogenetic analyses relating to early whales[5,6]. These authors based raoellid scores on *Khirtharia* and *Indohyus*. In the present analyses we have split scores for these animals, with *Khirtharia* scores based mostly on published H-GSP material and one unpublished skull (H-GSP 1979; dentition published[48], specimen now lost). Scores for *Indohyus* are based on the material in the RR and IITR-SB collections; all raoellid scores are listed, with the specimen number of the fossil on which the score was based, in Supplementary Table 1.

Study of tympanic walls. Tympanic wall thickness was investigated to address the presence of the involucrum quantitatively. The involucrum is the thickened medial wall of the tympanic bone (the ossified wall of the middle ear cavity). The lateral tympanic wall of cetaceans is reduced in thickness (the tympanic plate). To quantify these differences in tympanic walls, we calculated the ratio of medial tympanic wall thickness to the lateral tympanic wall thickness. Lateral tympanic wall thickness was measured with a micrometer (Dyer gauge) just inferior to the tympanic ring, and medial tympanic wall thickness was measured directly across from this site on the other (medial) side of the middle ear cavity (see Supplementary Table 2).

Bone histology. Bone histology was studied to investigate the presence of osteosclerosis. Osteosclerosis is the thickening of the cortical bone. It was quantified as the ratio of medullary cavity width divided by bone width in the mediolateral plane, because left and right cortical thickness plus medullary cavity thickness equals bone width (see Supplementary Table 3). Measurements were taken on the femur with callipers.

Fossil limb shaft fragments were embedded in Buehler low-viscosity epoxy resin and sectioned with a diamond saw. Sections were mounted on frosted glass slides by using epoxy resin. Mounted sections were then ground down and polished to a thickness of about 75 μm by using a precision grinder with 600, 800 and 1,200 grit paper[20] (Fig. 1i–l).

Study of stable isotopes. For the analysis of stable isotopes, three or more specimens of each species were analysed (when available; see Supplementary Tables 4 and 5) to provide a robust estimate of the population mean and s.d. for carbon and oxygen isotope values[33]. About 5 mg of enamel powder was collected from each specimen, either by drilling directly from the tooth or by grinding enamel chips in an agate mortar and pestle. Before collection, contaminants were removed by abrading the outer surface of the specimen.

Preparation of powders for analysis of stable isotopes followed published methods[31]. Powders were first transferred to 1-ml microcentrifuge vials and then soaked sequentially overnight in about 0.20 ml of a sodium hypochlorite solution (1–2 g dl^{-1}) and then in about 0.20 ml of calcium acetate buffered acetic acid (pH about 5.1). On addition of each reagent, samples were agitated for 1 min on a Vortex Genie vortex mixer. After each soak, the supernatant was removed by aspiration and the residual powder was rinsed five times with deionized water. Samples were then freeze-dried overnight and about 1.5 mg of powder from each was weighed into individual test tubes for analysis on a Thermo-Finnigan gas bench autosampler attached to a Thermo Finnigan Delta^{Plus} XP continuous-flow isotope-ratio mass spectrometer at the University of Wyoming Stable Isotope Facility.

All values for stable isotopes are reported in delta (δ) notation, using the equation $\delta(‰) = 1{,}000 \times (R_{sample}/R_{standard} - 1)$, where R_{sample} is the observed isotope ratio of the sample ($^{13}C/^{12}C$ or $^{18}O/^{16}O$) and $R_{standard}$ is the accepted ratio for an appropriate international standard (Vienna Pee Dee belemnite for $\delta^{13}C$; Vienna Standard Mean Ocean Water for $\delta^{18}O$). Analytical precision is typically better than 0.1‰ for $\delta^{13}C$ values and 0.2‰ for $\delta^{18}O$ values ($\pm 1\sigma$).

41. Ducrocq, S. The late Eocene Anthracotheriidae (Mammalia, Artiodactyla) from Thailand. *Palaeontographica A* **252**, 93–140 (1999).
42. Ducrocq, S. Unusual dental morphologies in late Eocene anthracotheres (Artiodactyla, Mammalia) from Thailand: dental anomalies and extreme variations. *N. Jb. Geol. Paläont. MH* **4**, 199–212 (1999).
43. Suteethorn, V., Buffetaut, E., Helmcke-Ingavat, R., Jaeger, J. J. & Jongkanjanasoontorn, Y. Oldest known Tertiary mammals from South East Asia: middle Eocene primate and anthracotheres from Thailand. *N. Jb. Geol. Paläont. MH* **9**, 563–570 (1988).
44. Colbert, E. H. Fossil mammals from Burma. *Am. Mus. Nat. Hist. Bull.* **74**, 419–424 (1938).
45. Brunet, M. M. Découverte d'un crâne d'Anthracotheriidae, *Microbunodon minimum* (Cuvier), á La Milloque (Lot-et-Garonne). *C. R. Acad. Sci. Paris* **267**, 835–838 (1968).
46. Lihoreau, F., Blondel, C., Barry, J. & Brunet, M. A new species of the genus *Microbunodon* (Anthracotheriidae, Artiodactyla) from the Miocene of Pakistan: genus revision, phylogenetic relationships and palaeobiogeography. *Zool. Scr.* **33**, 97–115 (2004).
47. Bajpai, S. *et al.* Early Eocene land mammals from Vastan Lignite Mine, District Surat, Gujarat, western India. *J. Palaeontol. Soc. India* **50**, 101–113 (2005).
48. West, R. M. Middle Eocene large mammal assemblage with Tethyan affinities, Ganda Kas region, Pakistan. *J. Paleontol.* **54**, 508–533 (1980).

Vol 450 | 20/27 December 2007 | doi:10.1038/nature06416

Locally dynamic synaptic learning rules in pyramidal neuron dendrites

Christopher D. Harvey[1,2] & Karel Svoboda[1,2]

Long-term potentiation (LTP) of synaptic transmission underlies aspects of learning and memory. LTP is input-specific at the level of individual synapses, but neural network models predict interactions between plasticity at nearby synapses. Here we show in mouse hippocampal pyramidal cells that LTP at individual synapses reduces the threshold for potentiation at neighbouring synapses. After input-specific LTP induction by two-photon glutamate uncaging or by synaptic stimulation, subthreshold stimuli, which by themselves were too weak to trigger LTP, caused robust LTP and spine enlargement at neighbouring spines. Furthermore, LTP induction broadened the presynaptic–postsynaptic spike interval for spike-timing-dependent LTP within a dendritic neighbourhood. The reduction in the threshold for LTP induction lasted ~10 min and spread over ~10 μm of dendrite. These local interactions between neighbouring synapses support clustered plasticity models of memory storage and could allow for the binding of behaviourally linked information on the same dendritic branch.

Long-lasting modifications of synaptic strength (LTP) are critical for learning and memory in many parts of the brain, including the hippocampus[1]. The extent to which LTP is synapse-specific influences the information processing and storage of a neuron. LTP can be input-specific[2], even at the level of individual synapses[3], indicating that synapses may function as independent units of plasticity[4]. However, neighbouring synapses might be co-regulated due to the heterosynaptic spread of LTP over short stretches of dendrite[5].

Neural network models predict interactions between plasticity at nearby synapses. Heterosynaptic metaplasticity suggests that LTP at one set of synapses may subsequently increase the threshold for potentiation at other synapses[6,7]. In contrast, clustered plasticity models[8–10] predict a decrease in the threshold for LTP in the neighbourhood of recently potentiated synapses, for example, owing to local synaptic tagging[10–12]. To distinguish between these possibilities, we probed the coupling between plasticity at nearby synapses using two-photon glutamate uncaging[3,13–16] combined with two-photon laser scanning microscopy[17,18] and perforated patch whole-cell recordings of synaptic currents.

Crosstalk between plasticity at nearby synapses

Does LTP at one synapse influence the threshold for plasticity at neighbouring synapses? We looked for such 'crosstalk' in acute hippocampal slices from green fluorescent protein (GFP)-expressing transgenic mice[19]. Dendritic spines were imaged on proximal (distance to the soma, <100 μm) secondary and tertiary apical dendrites of CA1 pyramidal neurons (Fig. 1a, c, e). Glutamate receptors on individual spines were stimulated with two-photon glutamate uncaging, and the resulting uncaging-evoked excitatory postsynaptic currents (uEPSCs) were measured at the soma using perforated patch-clamp recordings.

To induce LTP at individual spines, we paired a train of 30 stimuli (0.5 Hz) with postsynaptic depolarization to ~0 mV (ref. 3). In this 'LTP protocol', each uncaging stimulus (4 ms duration) triggered NMDA-R (*N*-methyl-D-aspartate receptor)-mediated spine $[Ca^{2+}]$ accumulations that were similar to $[Ca^{2+}]$ transients evoked by low-frequency synaptic stimulation at 0 mV (ref. 20) or by tetanic stimulation[21] (Supplementary Fig. 1b, c; see Supplementary Information). $[Ca^{2+}]$ accumulations were restricted to the stimulated spine (Supplementary Fig. 1a–c), indicating that glutamate did not spread to activate neighbouring spines. As a readout of plasticity, we monitored spine volumes and uEPSCs in response to test stimuli at the spine receiving the LTP protocol (LTP spine) and at neighbouring spines less than 4 μm from the LTP spine on the same branch. The LTP protocol resulted in a long-lasting (>40 min) increase in uEPSC amplitude and spine volume (Vol) at the LTP spine, but not at nearby spines ($\Delta uEPSC_{LTP\,spine} = 99 \pm 17\%$ (mean ± s.e.m.), $P < 0.01$; $\Delta uEPSC_{nearby\,spine} = -1 \pm 9\%$, $P > 0.9$; $\Delta Vol_{LTP\,spine} = 78 \pm 10\%$, $P < 0.01$; $\Delta Vol_{nearby\,spines} = 0 \pm 4\%$, $P > 0.9$; Fig. 1a, b). A similar protocol, but in which the amplitudes of NMDA-R-mediated spine $[Ca^{2+}]$ transients were reduced by a factor of four (subthreshold protocol, 1-ms pulse duration; Supplementary Fig. 1b, c), did not change uEPSC amplitude or spine volume at the spine receiving the uncaging stimuli (sub spine) or at nearby spines ($\Delta uEPSC_{sub\,spine} = -1 \pm 2\%$, $P > 0.4$; $\Delta uEPSC_{nearby\,spine} = 2 \pm 2\%$, $P > 0.6$; $\Delta Vol_{sub\,spine} = 1 \pm 1\%$, $P > 0.6$; $\Delta Vol_{nearby\,spines} = 1 \pm 4\%$, $P > 0.8$; Fig. 1c, d).

To test for crosstalk, we induced LTP at one spine (LTP spine) and, 90 s later, provided the subthreshold protocol at a neighbouring spine (sub spine). The subthreshold protocol now triggered LTP and a long-lasting spine enlargement ($\Delta uEPSC_{LTP\,spine} = 95 \pm 11\%$, $P < 0.01$; $\Delta uEPSC_{sub\,spine} = 97 \pm 10\%$, $P < 0.01$; $\Delta Vol_{LTP\,spine} = 76 \pm 16\%$, $P < 0.02$; $\Delta Vol_{sub\,spine} = 81 \pm 10\%$, $P < 0.01$; Fig. 1e, f). The levels of functional and structural plasticity were similar in spines receiving the LTP and subthreshold protocols (uEPSC, $P > 0.5$; Vol, $P > 0.5$; Fig. 1g). Other nearby spines that received neither stimulus did not change ($\Delta Vol = 1 \pm 1\%$, $P > 0.7$). Crosstalk did not occur after application of the LTP protocol at a postsynaptic potential of approximately −70 mV, which did not induce LTP, arguing that crosstalk is triggered by LTP induction and not by the uncaging process itself (see Supplementary Information). LTP induction at one spine therefore lowered the threshold for potentiation at nearby spines while maintaining input specificity.

[1]Janelia Farm Research Campus, HHMI, Ashburn, Virginia 20147, USA. [2]Watson School of Biological Sciences, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA.

The changes in uEPSC amplitude and spine volume were highly correlated[3,22] ($r = 0.86$, $P < 0.0001$; Fig. 1h), consistent with documented relationships between spine volume, postsynaptic density area and the number of AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptors in the postsynaptic density[14,23,24]. These observations confirm that spine enlargement is a structural correlate of LTP[3,22].

Crosstalk in unperturbed neurons

The pairing LTP protocol (Fig. 1) has non-physiological features. For example, depolarization of the postsynaptic neuron during pairing causes global Ca^{2+} influx through voltage-gated calcium channels, which could contribute to the crosstalk between synapses. To test if crosstalk occurs without sustained postsynaptic depolarization, we stimulated NMDA-Rs on individual spines from unperturbed neurons (in nominally 0 mM Mg^{2+}). Uncaging stimuli triggered [Ca^{2+}] transients that were restricted to the activated spine (Supplementary Fig. 1a–c). Each uncaging pulse during the LTP protocol produced NMDA-R currents (7.9 ± 1.1 pA; Supplementary Fig. 1a, d) that corresponded to the opening of ~5 NMDA-Rs, comparable to the number of receptors opened by low-frequency synaptic stimulation[25]. The LTP protocol triggered a large transient increase in spine volume in the LTP spine that decayed to a persistent spine enlargement after 10 min; spines neighbouring the stimulated spine did not change ($\Delta Vol_{LTP\ spine} = 76 \pm 18\%$, $P < 0.01$; $\Delta Vol_{nearby\ spines} = -1 \pm 4\%$, $P > 0.7$; Fig. 2b, e, f). The subthreshold protocol, which produced approximately fourfold lower NMDA-R currents and [Ca^{2+}] accumulations (Supplementary Fig. 1a–d), triggered only transient changes in spine volume that decayed within 10 min ($\Delta Vol_{sub\ spine} = 5 \pm 6\%$, $P > 0.3$; Fig. 2c, e, f). We next provided the LTP protocol at one spine and, 90 s later, tested for crosstalk by applying the subthreshold protocol at a neighbouring spine. The subthreshold protocol now induced sustained spine enlargement of the same size as that induced by the LTP protocol ($\Delta Vol_{LTP\ spine} = 66 \pm 8$ %, $P < 0.0001$; $\Delta Vol_{sub\ spine} = 67 \pm 10\%$, $P < 0.0001$; LTP spine versus sub spine, $P > 0.95$; Fig. 2d–f). Other spines that received neither stimulus did not change ($\Delta Vol = 0 \pm 1\%$, $P > 0.95$). Similar results were obtained in cultured rat hippocampal slices (Supplementary Fig. 2a–d). Persistent postsynaptic depolarization therefore was not required to observe the crosstalk in plasticity between synapses.

Crosstalk with synaptically induced plasticity

Glutamate released by uncaging may activate a distinct set of receptors compared to synaptically released glutamate. We therefore tested if crosstalk occurs after synaptically induced plasticity. Schaffer collateral axons were stimulated (120 pulses, 2 Hz) in low extracellular Mg^{2+} (refs 3 and 26). This 'synaptic LTP protocol' induced long-lasting spine enlargement in a sparse subset of spines (see Methods). The magnitude of the spine volume change ($\Delta Vol_{synaptic\ LTP\ spine} =$

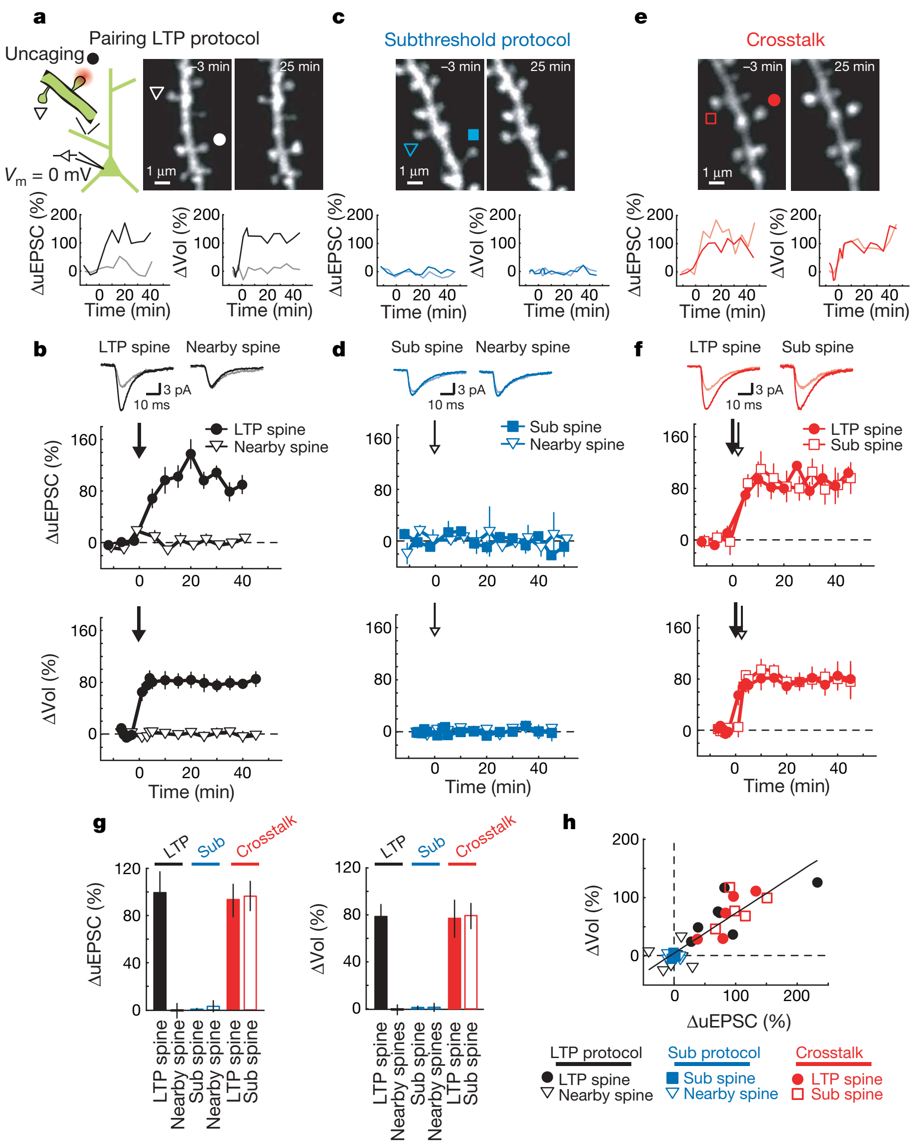


Figure 1 | Crosstalk with pairing-induced LTP. **a**, Top left, schematic of the experiment. V_m, membrane potential. Right, images before (−3 min) and after (25 min) LTP induction. At time = 0 the LTP protocol (30 uncaging pulses at 0.5 Hz, 4-ms pulse duration, postsynaptic potential 0 mV) was applied to the spine marked by a circle (LTP spine). A triangle marks a tested nearby spine. Lower panels show changes in uEPSC amplitude and spine volume at the LTP (black) and nearby (grey) spines. **b**, Upper panels, uEPSCs, averaged across all cells, in response to test stimuli before (−3 min; grey) and after (40 min; black) the LTP protocol. Lower panels, time course of the changes in uEPSC amplitude and spine volume at the LTP spine (filled circles; $n = 7$) and at nearby spines (open triangles; uEPSC, $n = 7$; Vol, $n = 31$). The arrow marks the LTP protocol. **c**, **d**, Same as for **a** and **b** except with the subthreshold protocol. At time = 0 the subthreshold protocol (30 uncaging pulses at 0.5 Hz, 1-ms pulse duration, postsynaptic potential 0 mV) was applied to the spine marked by a filled square (sub spine; $n = 5$). Open triangles indicate nearby spines (uEPSC, $n = 5$; Vol, $n = 26$). **e**, **f**, Same as for **a** and **b**, except for the crosstalk case. At time = 0 the LTP protocol was applied to the spine marked by a filled circle (LTP spine) and, 90 s later, the subthreshold protocol was given at the spine marked by an open square (sub spine). $n = 5$, mean ± s.e.m. **g**, Changes in uEPSC amplitude and spine volume. Error bars indicate mean ± s.e.m. **h**, Correlation between changes in uEPSC amplitude and spine volume. $r = 0.86$, $P < 0.0001$.

$70 \pm 14\%$, $P < 0.001$) was similar to that triggered by the uncaging LTP protocol[3] ($P > 0.8$; compare Fig. 3e and Fig. 2b). Spine enlargement was thus used to identify synapses potentiated by synaptic stimulation (see Methods). To test for crosstalk, we provided the subthreshold protocol at a nearby spine (sub spine) two minutes after the synaptic LTP protocol. The subthreshold protocol, which by itself did not trigger structural plasticity (Fig. 2c, e, f), now induced a persistent spine enlargement ($\Delta Vol_{sub\ spine} = 62 \pm 9\%$, $P < 0.001$) of similar magnitude to the synaptically induced volume change ($P > 0.6$; Fig. 3b–e). Other nearby spines did not change ($\Delta Vol_{nearby\ spines} = -3 \pm 5\%$, $P > 0.4$; Fig. 3b–e). Synaptically induced plasticity therefore reduced the threshold for potentiation at neighbouring synapses.

Modulation of the window for STDP

Excitatory postsynaptic potentials (EPSPs) followed by action potentials within a short time window (tens of milliseconds) can trigger LTP[27]. The magnitude of this spike-timing-dependent potentiation (STDP) decreases monotonically with the time between the EPSP and the action potential[28,29]. Because crosstalk reduces the threshold for potentiation in the neighbourhood of the LTP spine, crosstalk could broaden the spike time window (Δt) for STDP at neighbouring spines. We induced STDP with uncaging pulses (60 pulses, 2 Hz) followed ($\Delta t = 5$ ms) by three action potentials at 50 Hz. The amplitudes of uEPSPs (0.41 ± 0.19 mV, mean $\pm$ s.d.) were similar to those of miniature EPSPs[30]. This induction protocol induced long-lasting increases in the uEPSC amplitude and spine volume at the stimulated spine, but not at nearby spines within 4 µm on the same dendritic branch ($\Delta uEPSC_{\Delta t = 5\ ms} = 62 \pm 17\%$, $P < 0.02$; $\Delta uEPSC_{nearby\ spine} = 5 \pm 8\%$, $P > 0.5$; $\Delta Vol_{\Delta t = 5\ ms} = 57 \pm 13\%$, $P < 0.01$; $\Delta Vol_{nearby\ spines} = 0 \pm 3\%$, $P > 0.8$; Fig. 4b, d). The magnitudes of functional and structural plasticity decreased as the time between the uEPSP and the action potentials increased ($\tau_{\Delta uEPSC} = 17.6$ ms; $\tau_{\Delta Vol} = 16.6$ ms; Fig. 4c). Pairing at longer intervals ($\Delta t = 35$ ms) did not trigger LTP or spine enlargement ($\Delta uEPSC_{\Delta t = 35\ ms} = -3 \pm 10\%$, $P > 0.8$; $\Delta Vol_{\Delta t = 35\ ms} = 4 \pm 3\%$, $P > 0.2$; Fig. 4e), indicating that uEPSPs or action potentials alone were not sufficient to trigger LTP. STDP therefore was induced at single spines in an input-specific manner.

We next induced STDP at one spine with an uEPSP-to-action-potential time window of 5 ms, and, 90 s later, stimulated a

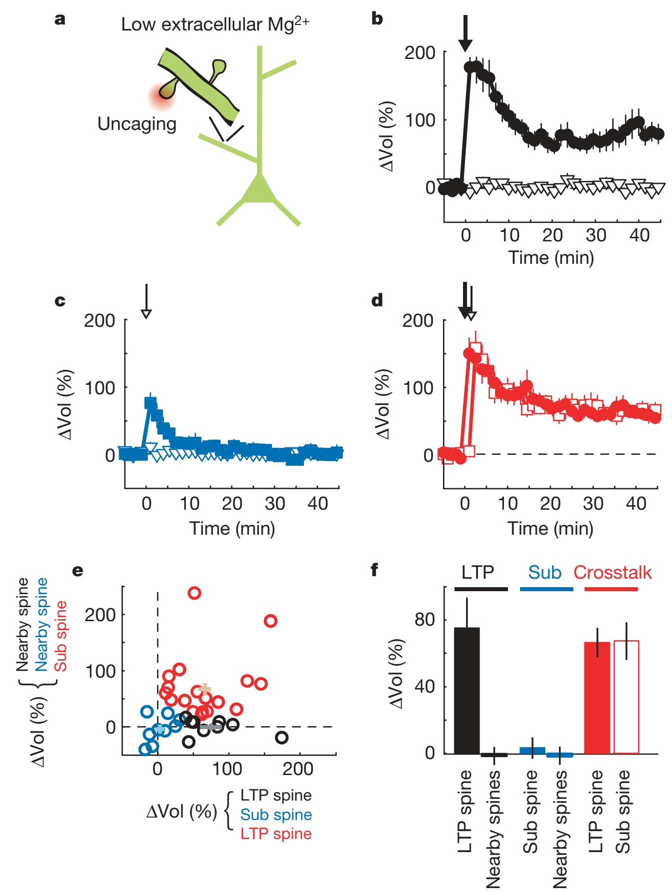


Figure 2 | Crosstalk in unperturbed neurons. **a**, Schematic of the experiment. **b**, Time course of the spine-volume changes induced by the LTP protocol (applied at time = 0, 30 uncaging pulses at 0.5 Hz, 4-ms pulse duration, in low extracellular Mg^{2+}) for the stimulated spine (LTP spine, closed circles; $n = 9$) and nearby spines (open triangle; $n = 29$). **c**, Time course of the spine-volume changes induced by the subthreshold protocol (applied at time = 0, 30 uncaging pulses at 0.5 Hz, 1-ms pulse duration, in low extracellular Mg^{2+}) for the stimulated spine (sub spine, filled squares; $n = 8$) and nearby spines (open triangles; $n = 38$). **d**, Time course of the spine-volume changes for the crosstalk case. At time = 0 the LTP protocol was applied to the LTP spine (filled circles) and, 90 s later, the subthreshold protocol was given at a neighbouring spine (sub spine, open squares). $n = 18$, mean $\pm$ s.e.m. **e**, Spine volume changes from individual experiments. Black, LTP protocol; blue, sub protocol; red, crosstalk. Crosses indicate mean $\pm$ s.e.m. **f**, Changes in spine volume. Error bars indicate mean $\pm$ s.e.m.

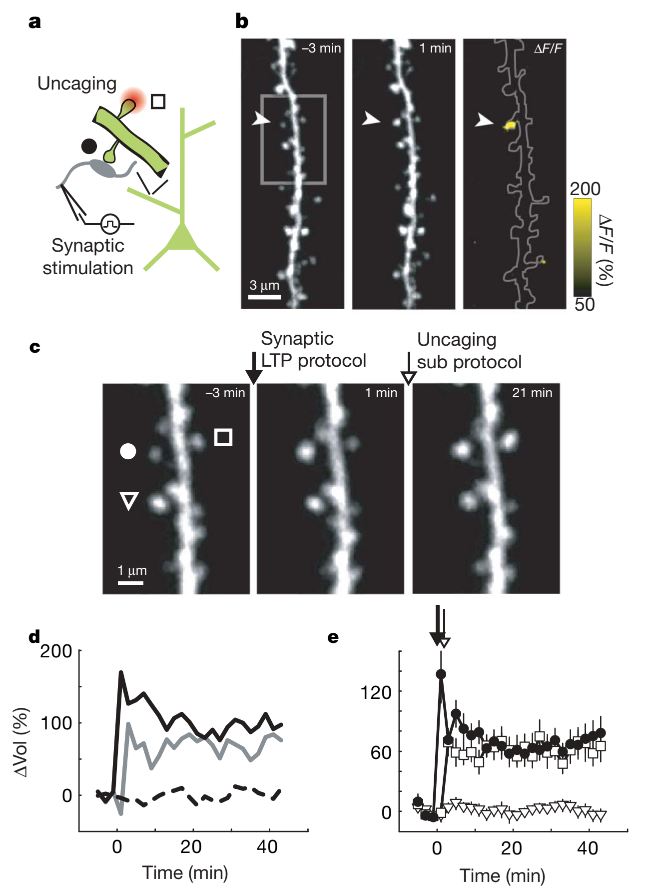


Figure 3 | Crosstalk with synaptically induced plasticity. **a**, Schematic of the experiment. **b**, Images before (-3 min) and after (1 min) the synaptic LTP protocol (applied at time = 0; 120 stimuli, 2 Hz in low extracellular Mg^{2+}). The arrowheads mark an enlarged spine (synaptic LTP spine). A ratio image ($\Delta F/F$) comparing fluorescence intensity before (-3 min) and after (1 min) the synaptic LTP protocol is shown. **c**, High magnification images before stimulation (-3 min), after the synaptic LTP protocol (1 min; applied at time = 0), and after the subthreshold protocol (21 min; applied at time = 2 min). The circle, square and triangle mark the synaptic LTP spine, the sub spine and a nearby spine, respectively. **d**, Changes in spine volume for the example shown in **b** and **c**. Black solid line, synaptic LTP spine; grey solid line, sub spine; black dashed line, nearby spine. **e**, Time course of the change in spine volume for synaptic LTP spines (filled circle; $n = 11$), sub spines (open square; $n = 11$) and nearby spines (open triangle; $n = 34$), mean $\pm$ s.e.m.

neighbouring spine with an uEPSP–action potential interval of 35 ms. Under these conditions, the uEPSP–action potential pairing at the 35-ms time window now induced LTP and a long-lasting spine enlargement ($\Delta uEPSC_{\Delta t = 5\,ms} = 67 \pm 10\%$, $P < 0.01$; $\Delta uEPSC_{\Delta t = 35\,ms} = 69 \pm 8\%$, $P < 0.01$; $\Delta Vol_{\Delta t = 5\,ms} = 68 \pm 9\%$, $P < 0.01$; $\Delta Vol_{\Delta t = 35\,ms} = 74 \pm 15\%$, $P < 0.02$; Fig. 4f). The levels of functional and structural plasticity were similar in spines receiving the pairing at short and long intervals (uEPSC, $P > 0.4$; Vol, $P > 0.4$; Fig. 4g), and the changes in uEPSC amplitude and spine volume were highly correlated ($r = 0.81$, $P < 0.0001$; Fig. 4h). Other nearby spines that received neither stimulus did not change ($\Delta Vol = -1 \pm 1\%$, $P > 0.7$). LTP induction at one spine therefore broadened the uEPSP–action potential time window for STDP at neighbouring spines.

Characterization of crosstalk

We next measured the timescale of the crosstalk in plasticity between synapses. We varied the time between the LTP and subthreshold protocols given in low extracellular Mg^{2+} while maintaining the distance between the stimulated spines at ~3 μm. The crosstalk was measured as the volume change triggered by the subthreshold protocol at the sub spine after LTP induction at the LTP spine. Crosstalk decreased gradually with time and lasted for up to 10 min ($t_{1/2} = 5.3$ min; Fig. 5a).

To determine the length scale of the crosstalk, we varied the distance between the spines receiving the LTP and subthreshold protocols while keeping the time between stimuli at 90 s. Crosstalk decreased gradually with distance for up to ~8 μm in both directions along the parent dendrite (full-width at half-maximum = 11.1 μm; Fig. 5b). The magnitude of crosstalk was similar for spines farther or closer to the apical trunk with respect to the spine receiving the LTP protocol (data not shown). The length scale of the crosstalk was similar in cultured rat hippocampal slices (full-width at half-maximum = 10.2 μm; Supplementary Fig. 2e). Consistently, when spines separated by ~10 μm were stimulated by the LTP and subthreshold protocols paired with depolarization to ~0 mV, the subthreshold protocol did not induce functional or structural plasticity ($\Delta uEPSC_{sub\,spine} = -7 \pm 5\%$, $P > 0.15$; $\Delta Vol_{sub\,spine} = -8 \pm 8\%$, $P > 0.4$; Fig. 5c). Furthermore, after synaptically induced spine enlargement, the subthreshold protocol did not trigger structural plasticity at spines located ~10 μm from the enlarged spine ($\Delta Vol_{sub\,spine} = -3 \pm 8\%$, $P > 0.9$; Fig. 5d).

Our experiments indicate that LTP induction activates a factor at the LTP spine that spreads to reduce the threshold for potentiation at neighbouring synapses. Extracellular diffusible factors have been implicated in the heterosynaptic spread of LTP[31,32]. Similarly, intracellular factors can spread over the relevant time and length scales[33,34] (C.D.H., Ryohei Yasuda and K.S., unpublished). To distinguish between extracellular and intracellular factors, we examined whether crosstalk can occur between spines that are close within the neuropil (<4 μm) but are located on different dendritic branches and therefore are far apart in terms of cytoplasmic distance (>50 μm). We induced LTP at one spine and, 90 s later, provided the subthreshold protocol at the sub spine less than 4 μm away on a nearby dendritic branch from the same cell. Under these conditions, the subthreshold

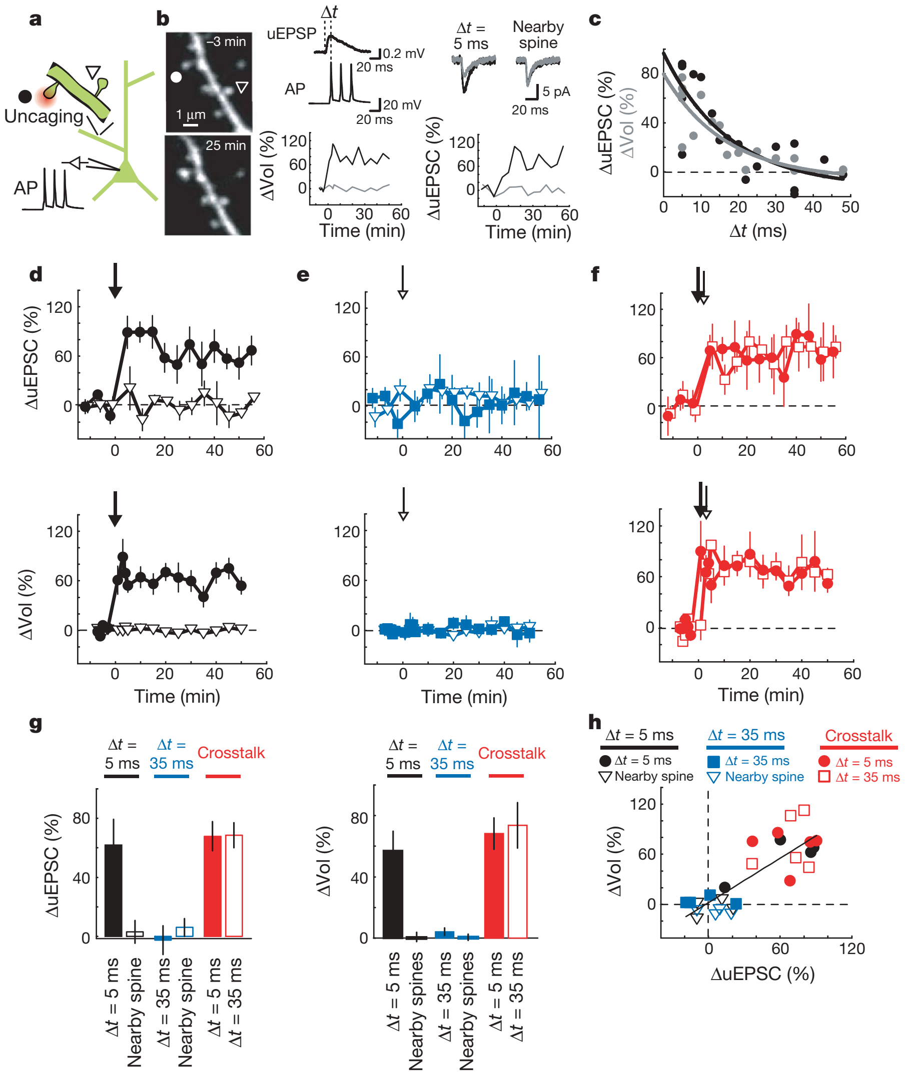


Figure 4 | Crosstalk with spike-timing-dependent LTP. **a**, Schematic of the experiment. **b**, Left, images before (−3 min) and after (25 min) spike-timing-dependent LTP induction. At time = 0, uncaging pulses (60 pulses at 2 Hz) followed by three action potentials at 50 Hz ($\Delta t = 5$ ms) were applied to the spine marked by the circle. The triangle marks a tested nearby spine. Top middle, example uEPSPs and action potentials (APs) from unpaired stimuli. Top right, uEPSCs averaged over 5 trials before (−6 min, grey line) and after (25 min, black line) uEPSP–action potential pairing. Bottom middle and bottom right, changes in uEPSC amplitude and spine volume at the stimulated (black) and nearby (grey) spines. **c**, Changes in uEPSC amplitude (black) and spine volume (grey) at different uEPSP–action potential time windows (Δt). Changes were measured from 20–30 min post stimulus. Exponential fits are shown. **d**, Time course of the changes in uEPSC amplitude and spine volume for uEPSP–action potential pairing at $\Delta t = 5$ ms (filled circle; $n = 4$) and at nearby spines (open triangle; uEPSC, $n = 4$; Vol, $n = 20$). The arrow marks the time of uEPSP–action potential pairing. **e**, Time course of the changes in uEPSC amplitude and spine volume for uEPSP–action potential pairing at $\Delta t = 35$ ms (filled square; $n = 4$) and at nearby spines (open triangle; uEPSC, $n = 4$; Vol, $n = 21$). **f**, Time course of the changes in uEPSC amplitude and spine volume for the crosstalk case. At time = 0, one spine was stimulated with uEPSP–action potential pairing at $\Delta t = 5$ ms (filled circle) and, 90 s later, a neighbouring spine was stimulated with uEPSP–action potential pairing at $\Delta t = 35$ ms (open square). $n = 5$, mean ± s.e.m. **g**, Changes in uEPSC amplitude and spine volume. Error bars indicate mean ± s.e.m. **h**, Correlation between changes in uEPSC amplitude and spine volume. $r = 0.81$, $P < 0.0001$.

protocol failed to induce structural plasticity ($\Delta Vol_{sub\ spine} = 1 \pm 9\%$, $P > 0.6$; Fig. 6a), indicating that intracellular factors, rather than extracellular factors, were necessary for the crosstalk between synapses.

Ca^{2+} release from intracellular stores has been implicated in the heterosynaptic spread of some forms of synaptic plasticity[35,36]. However, eliminating Ca^{2+} release from intracellular stores using thapsigargin (1 μM) and ryanodine (20 μM) (Supplementary Fig. 3) did not affect the crosstalk between synapses ($\Delta Vol_{sub\ spine} = 67 \pm 19\%$, $P > 0.95$) (Fig. 6b).

The crosstalk in plasticity between neighbouring synapses described here shares characteristics with synaptic tagging, in which early LTP at one set of synapses can be converted into late LTP by the strong stimulation of a second group of synapses[11]. Synaptic-tagging-based plasticity occurs both when the 'weak' stimulus precedes and when it follows the 'strong' stimulus[37,38]. We therefore tested if the crosstalk in plasticity depends on the order of stimuli. When the subthreshold protocol preceded the LTP protocol by 90 s, the subthreshold protocol did not induce spine enlargement ($\Delta Vol_{sub\ spine} = 2 \pm 14\%$, $P > 0.8$; Fig. 5a). Because synaptic-tagging-based crosstalk requires the capture of newly synthesized proteins[11,39], we also tested the role of protein synthesis in the crosstalk between neighbouring synapses. Application of the protein synthesis inhibitor anisomycin (25 μM) had no effect on the spine enlargement induced by the LTP and subthreshold protocols ($\Delta Vol_{LTP\ spine} = 63 \pm 11\%$, $P > 0.7$; $\Delta Vol_{sub\ spine} = 79 \pm 17\%$, $P > 0.3$; Fig. 6b). Similar results were obtained with other protein synthesis inhibitors (60 μM cycloheximide, $\Delta Vol_{sub\ spine} = 64 \pm 16\%$, $P > 0.9$; 50 μM emetine, $\Delta Vol_{sub\ spine} = 78 \pm 8\%$, $P > 0.6$). As a positive control for inhibitor function, anisomycin, cycloheximide and emetine caused a rapid decrease in destabilized EGFP fluorescence[40] (Supplementary Fig. 4). The crosstalk in plasticity between neighbouring spines is therefore distinct from synaptic tagging.

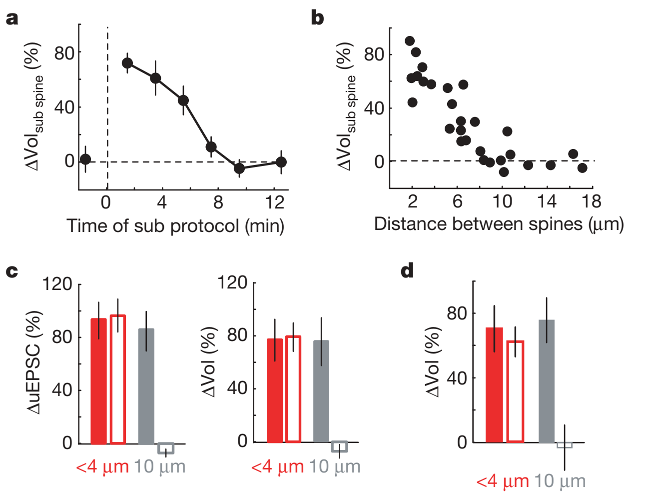


Figure 5 | Spatial and temporal scales of crosstalk. **a**, Timescale of crosstalk. At time = 0, the LTP protocol was applied to a single spine (in low extracellular Mg^{2+}, see Fig. 2). The subthreshold protocol was applied to a neighbouring spine (sub spine) ~3 μm away. $n \geq 4$ for all time points, mean ± s.e.m. **b**, Length scale of crosstalk. The LTP protocol was applied to a single spine and, 90 s later, a nearby spine (sub spine) was stimulated with the subthreshold protocol (in low extracellular Mg^{2+}, see Fig. 2). **c**, Distance-dependence of crosstalk with pairing-induced plasticity. The LTP protocol (at postsynaptic potential 0 mV) was applied to the LTP spine and, 90 s later, the subthreshold protocol (at postsynaptic potential 0 mV) was applied to a spine (sub spine) either less than 4 μm or ~10 μm away (see Fig. 1). The data for spines separated by less than 4 μm are from Fig. 1. $n = 4$ at 10 μm, mean ± s.e.m. Filled bars, LTP spine; open bars, sub spine. **d**, Distance-dependence of crosstalk with synaptically induced plasticity. The synaptic LTP protocol was applied in low extracellular Mg^{2+} to induce plasticity in the synaptic LTP spine. Two minutes later, the subthreshold protocol was applied to a spine (sub spine) either less than 4 μm or ~10 μm away (see Fig. 3). The data for spines separated by less than 4 μm are from Fig. 3. $n = 5$ at 10 μm, mean ± s.e.m. Filled bars, synaptic LTP spine; open bars, sub spine.

Discussion

We have shown that the induction of plasticity at individual synapses can be influenced by events at neighbouring synapses. LTP induction at one synapse decreased the threshold for potentiation at nearby synapses within ~10 μm for ~10 min. Crosstalk did not perturb input-specificity *per se*, and therefore differed from the heterosynaptic spread of plasticity[5,31,32,35]. However, the reduction in LTP induction threshold in the vicinity of a potentiated synapse may help explain discrepancies between data showing the heterosynaptic spread of LTP[5] and synapse-specific LTP at single spines[3]. Previous studies have suggested that synaptic plasticity can be influenced by prior neural activity[6,7,11,12,41–43]. However, the time courses of these interactions were much longer than the timescale of crosstalk reported here. Furthermore, these studies did not establish a length scale for crosstalk.

What cellular mechanisms could underlie the crosstalk in plasticity between neighbouring synapses? Our results indicate that the intersynaptic spread of intracellular signalling factors probably has a key role. The timescale and spatial scale of crosstalk are consistent with a diffusing cytoplasmic factor[33,34]. This factor could modify synaptic properties at nearby spines to decrease the threshold for LTP or may provide enzymatic activity that is necessary for LTP induction but is not produced by subthreshold stimuli.

Although synaptic modifications can occur in an input-specific manner[3] (Figs 1a, b, 2b and 4d), the coordinated regulation of groups of 10–20 synapses within a dendritic neighbourhood indicates that individual synapses do not necessarily function as independent units of plasticity. Models of clustered plasticity[8–10] propose that individual engrams could be stored in synapses sharing the same dendritic branch, which would increase the information storage capacity of the neuron through the nonlinear summation of synaptic inputs[8,9,30,44]. Clustered plasticity implies the binding of inputs that are active during the same behavioural epochs on the same dendritic branch. It will be of interest to map the distribution of the information carried by synapses within the dendritic trees of individual neurons.

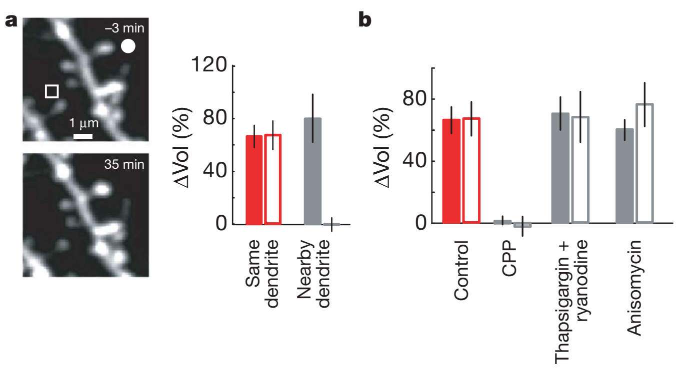


Figure 6 | Signalling underlying crosstalk. **a**, Crosstalk examined at nearby dendritic branches. The LTP protocol was applied to the spine marked by a circle (LTP spine) and, 90 s later, the subthreshold protocol was given at a spine on a nearby dendritic branch from the same cell (sub spine, square; in low extracellular Mg^{2+}, see Fig. 2). Example images are shown. The 'same dendrite' data are from Fig. 2. $n = 5$ for nearby dendrites, mean ± s.e.m. Filled bars, LTP spine; open bars, sub spine. **b**, Pharmacological analysis of crosstalk. In the presence of the specified drugs, the LTP protocol was applied to the LTP spine and, 90 s later, a neighbouring spine (sub spine) was stimulated with the subthreshold protocol (in low extracellular Mg^{2+}, see Fig. 2). Control data are from Fig. 2. $n = 4$ each for CPP (3-(*R*-2-carboxypiperazin-4-yl)-propyl-1-phosphonic acid; 10 μM), thapsigargin (1 μM) + ryanodine (20 μM), and anisomycin (25 μM), mean ± s.e.m. Filled bars, LTP spine; open bars, sub spine.

METHODS SUMMARY

Acute hippocampal brain slices were prepared from Thy1 GFP mice[19] (line M; postnatal day 14–18). Two-photon laser-scanning microscopy and two-photon glutamate uncaging were performed using a custom-built microscope with two Ti:sapphire lasers. For glutamate uncaging, brief (1 or 4 ms) laser exposures were delivered ~0.5 μm from the tip of the spine head in the presence of 2.5 mM methoxy-nitroindoline(MNI)-caged-L-glutamate. Uncaging-evoked EPSCs (uEPSCs) were measured using amphotericin-mediated perforated patch-clamp recordings. For synaptic stimulation (Fig. 3), short current pulses (0.1 ms, 30 μA) were delivered with a glass pipette positioned close (10 to 20 μm) to a GFP-labelled dendrite of interest. Plasticity was induced using four protocols: depolarization to ~0 mV paired with 30 uncaging pulses at 0.5 Hz (Fig. 1); 30 uncaging pulses at 0.5 Hz in low extracellular Mg^{2+} (Fig. 2); 120 synaptic stimuli at 2 Hz in low extracellular Mg^{2+} (Fig. 3); and an uncaging pulse followed by three action potentials at 50 Hz, repeated 60 times at 2 Hz (Fig. 4). Spine volumes were measured as the integrated green fluorescence after background subtraction, which is proportional to spine volume[45], normalized to the fluorescence intensity of the thick apical dendrite[25].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 August; accepted 29 October 2007.

1. Malenka, R. C. & Bear, M. F. LTP and LTD: an embarrassment of riches. *Neuron* **44,** 5–21 (2004).
2. Andersen, P., Sundberg, S. H., Sveen, O. & Wigstrom, H. Specific long-lasting potentiation of synaptic transmission in hippocampal slices. *Nature* **266,** 736–737 (1977).
3. Matsuzaki, M., Honkura, N., Ellis-Davies, G. C. & Kasai, H. Structural basis of long-term potentiation in single dendritic spines. *Nature* **429,** 761–766 (2004).
4. Yuste, R. & Denk, W. Dendritic spines as basic functional units of neuronal integration. *Nature* **375,** 682–684 (1995).
5. Engert, F. & Bonhoeffer, T. Synapse specificity of long-term potentiation breaks down at short distances. *Nature* **388,** 279–284 (1997).
6. Abraham, W. C., Mason-Parker, S. E., Bear, M. F., Webb, S. & Tate, W. P. Heterosynaptic metaplasticity in the hippocampus *in vivo*: a BCM-like modifiable threshold for LTP. *Proc. Natl Acad. Sci. USA* **98,** 10924–10929 (2001).
7. Wang, H. & Wagner, J. J. Priming-induced shift in synaptic plasticity in the rat hippocampus. *J. Neurophysiol.* **82,** 2024–2028 (1999).
8. Poirazi, P. & Mel, B. W. Impact of active dendrites and structural plasticity on the memory capacity of neural tissue. *Neuron* **29,** 779–796 (2001).
9. Mehta, M. R. Cooperative LTP can map memory sequences on dendritic branches. *Trends Neurosci.* **27,** 69–72 (2004).
10. Govindarajan, A., Kelleher, R. J. & Tonegawa, S. A clustered plasticity model of long-term memory engrams. *Nature Rev. Neurosci.* **7,** 575–583 (2006).
11. Frey, U. & Morris, R. G. Synaptic tagging and long-term potentiation. *Nature* **385,** 533–536 (1997).
12. Martin, K. C. *et al.* Synapse-specific, long-term facilitation of aplysia sensory to motor synapses: a function for local protein synthesis in memory storage. *Cell* **91,** 927–938 (1997).
13. Furuta, T. *et al.* Brominated 7-hydroxycoumarin-4-ylmethyls: photolabile protecting groups with biologically useful cross-sections for two photon photolysis. *Proc. Natl Acad. Sci. USA* **96,** 1193–1200 (1999).
14. Matsuzaki, M. *et al.* Dendritic spine geometry is critical for AMPA receptor expression in hippocampal CA1 pyramidal neurons. *Nature Neurosci.* **4,** 1086–1092 (2001).
15. Carter, A. G. & Sabatini, B. L. State-dependent calcium signaling in dendritic spines of striatal medium spiny neurons. *Neuron* **44,** 483–493 (2004).
16. Sobczyk, A., Scheuss, V. & Svoboda, K. NMDA receptor subunit-dependent $[Ca^{2+}]$ signaling in individual hippocampal dendritic spines. *J. Neurosci.* **25,** 6037–6046 (2005).
17. Denk, W., Strickler, J. H. & Webb, W. W. Two-photon laser scanning microscopy. *Science* **248,** 73–76 (1990).
18. Svoboda, K. & Yasuda, R. Principles of two-photon excitation microscopy and its applications to neuroscience. *Neuron* **50,** 823–839 (2006).
19. Feng, G. *et al.* Imaging neuronal subsets in transgenic mice expressing multiple spectral variants of GFP. *Neuron* **28,** 41–51 (2000).
20. Sabatini, B. S., Oertner, T. G. & Svoboda, K. The life-cycle of Ca^{2+} ions in spines. *Neuron* **33,** 439–452 (2002).
21. Muller, W. & Connor, J. A. Dendritic spines as individual neuronal compartments for synaptic Ca^{2+} responses. *Nature* **354,** 73–76 (1991).
22. Kopec, C. D., Li, B., Wei, W., Boehm, J. & Malinow, R. Glutamate receptor exocytosis and spine enlargement during chemically induced long-term potentiation. *J. Neurosci.* **26,** 2000–2009 (2006).
23. Nusser, Z. *et al.* Cell type and pathway dependence of synaptic AMPA receptor number and variability in the hippocampus. *Neuron* **21,** 545–559 (1998).
24. Takumi, Y., Ramirez-Leon, V., Laake, P., Rinvik, E. & Ottersen, O. P. Different modes of expression of AMPA and NMDA receptors in hippocampal synapses. *Nature Neurosci.* **2,** 618–624 (1999).
25. Nimchinsky, E. A., Yasuda, R., Oertner, T. G. & Svoboda, K. The number of glutamate receptors opened by synaptic stimulation in single hippocampal spines. *J. Neurosci.* **24,** 2054–2064 (2004).
26. Lang, C. *et al.* Transient expansion of synaptically connected dendritic spines upon induction of hippocampal long-term potentiation. *Proc. Natl Acad. Sci. USA* **101,** 16665–16670 (2004).
27. Dan, Y. & Poo, M. M. Spike timing-dependent plasticity of neural circuits. *Neuron* **44,** 23–30 (2004).
28. Bi, G. Q. & Poo, M. M. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.* **18,** 10464–10472 (1998).
29. Wittenberg, G. M. & Wang, S. S. Malleability of spike-timing-dependent plasticity at the CA3–CA1 synapse. *J. Neurosci.* **26,** 6610–6617 (2006).
30. Losonczy, A. & Magee, J. C. Integrative properties of radial oblique dendrites in hippocampal CA1 pyramidal neurons. *Neuron* **50,** 291–307 (2006).
31. Scanziani, M., Malenka, R. C. & Nicoll, R. A. Role of intercellular interactions in heterosynaptic long-term depression. *Nature* **380,** 446–450 (1996).
32. Schuman, E. M. & Madison, D. V. A requirement for the intercellular messenger nitric oxide in long-term potentiation. *Science* **254,** 1503–1506 (1991).
33. Gray, N. W., Weimer, R. M., Bureau, I. & Svoboda, K. Rapid Redistribution of synaptic PSD-95 in the neocortex *in vivo. PLoS Biol.* **4,** e370 (2006).
34. Tsuriel, S. *et al.* Local sharing as a predominant determinant of synaptic matrix molecular dynamics. *PLoS Biol.* **4,** e271 (2006).
35. Nishiyama, M., Hong, K., Mikoshiba, K., Poo, M. M. & Kato, K. Calcium stores regulate the polarity and input specificity of synaptic modification. *Nature* **408,** 584–588 (2000).
36. Royer, S. & Pare, D. Conservation of total synaptic weight through balanced synaptic depression and potentiation. *Nature* **422,** 518–522 (2003).
37. Frey, U. & Morris, R. G. Weak before strong: dissociating synaptic tagging and plasticity-factor accounts of late-LTP. *Neuropharmacology* **37,** 545–552 (1998).
38. Casadio, A. *et al.* A transient, neuron-wide form of CREB-mediated long-term facilitation can be stabilized at specific synapses by local protein synthesis. *Cell* **99,** 221–237 (1999).
39. Fonseca, R., Nagerl, U. V., Morris, R. G. & Bonhoeffer, T. Competing for memory: hippocampal LTP under regimes of reduced protein synthesis. *Neuron* **44,** 1011–1020 (2004).
40. Li, X. *et al.* Generation of destabilized green fluorescent protein as a transcription reporter. *J. Biol. Chem.* **273,** 34970–34975 (1998).
41. Abraham, W. C. & Bear, M. F. Metaplasticity: the plasticity of synaptic plasticity. *Trends Neurosci.* **19,** 126–130 (1996).
42. Turrigiano, G. G. & Nelson, S. B. Homeostatic plasticity in the developing nervous system. *Nature Rev. Neurosci.* **5,** 97–107 (2004).
43. Huang, Y. Y., Colino, A., Selig, D. K. & Malenka, R. C. The influence of prior synaptic activity on the induction of long-term potentiation. *Science* **255,** 730–733 (1992).
44. Golding, N. L., Staff, N. P. & Spruston, N. Dendritic spikes as a mechanism for cooperative long-term potentiation. *Nature* **418,** 326–331 (2002).
45. Holtmaat, A. J. *et al.* Transient and persistent dendritic spines in the neocortex *in vivo. Neuron* **45,** 279–291 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. Zhong and R. Yasuda for discussions, T. O'Connor for programming assistance, K. H. Wang for *destabilized EGFP* DNA, and R. Malinow and J. Magee for comments on the manuscript. This work was supported by HHMI, by the NIH, and by a David and Fanny Luke Fellowship (C.D.H.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to K.S. (svobodak@janelia.hhmi.org).

METHODS

Preparation. Acute hippocampal brain slices (300 μm thick) from Thy1 GFP mice[19] (line M; postnatal day 14–18) were prepared in accordance with the animal care and use guidelines of Cold Spring Harbor Laboratory and Janelia Farm Research Campus. Slices were cut in gassed (95% O_2/5% CO_2), ice-cold cutting solution containing 110 mM choline chloride, 25 mM $NaHCO_3$, 25 mM D-glucose, 2.5 mM KCl, 7 mM $MgCl_2$, 0.5 mM $CaCl_2$, 1.25 mM NaH_2PO_4, 11.5 mM sodium ascorbate and 3 mM sodium pyruvate. Slices were then incubated in gassed artificial cerebral spinal fluid (ACSF) containing 127 mM NaCl, 25 mM $NaHCO_3$, 25 mM D-glucose, 2.5 mM KCl, 1 mM $MgCl_2$, 2 mM $CaCl_2$ and 1.25 mM NaH_2PO_4 at 35 °C for 30 min and then at room temperature (22–24 °C) until used.

Hippocampal slice cultures (Supplementary Figs 2 and 4) were prepared from postnatal day 6 or 7 rats[46], in accordance with institutional animal care and use guidelines. After 5–8 days in culture, cells were transfected by ballistic gene transfer using gold beads (~15 mg, 1.6 μm diameter) coated with 10 μg of plasmid DNA. Experiments were performed 2–3 days post-transfection.

Experiments were performed at room temperature except for those in Fig. 3 (33 °C). MNI-caged-L-glutamate, CPP, NBQX (2,3-dioxo-6-nitro-1,2,3,4-tetrahydrobenzo[f]quinoxaline-7-sulfonamide), thapsigargin and ryanodine were from Tocris; amphotericin B was from Sigma; and TTX (tetrodotoxin), anisomycin, emetine and cycloheximide were from Calbiochem.

Electrophysiology. Perforated patch-clamp recordings were used to prevent the washout of intracellular signalling molecules and LTP[28,47]. The internal solution contained 136.5 mM potassium gluconate, 17.5 mM KCl, 9 mM NaCl, 1 mM $MgCl_2$, 10 mM HEPES, 0.2 mM EGTA and 0.5 mg ml^{-1} amphotericin B. Pipettes were front-filled with a small volume of internal solution without amphotericin B. Perforations reached a stable series resistance (36 ± 8 MΩ, mean ± s.d.) within 30–45 min of seal formation. Series resistances were stable (±20%) throughout the experiment. uEPSCs were measured in response to test stimuli (0.1 Hz) at −70 mV. uEPSC amplitudes were measured as the difference between the mean current amplitude over a 5-ms window around the peak and the mean current amplitude over a 100-ms window before the uncaging stimulus. Each time point is the average of five trials (Fig. 1b, d, f and Fig. 4d–f). Spike-timing-dependent LTP (Fig. 4) was induced in current-clamp mode. Action potentials were triggered by brief current injections at the soma (2 ms, 1–3 nA). Voltage-clamp whole-cell recordings for $[Ca^{2+}]$ imaging (Supplementary Fig. 1) were made using an internal solution containing 135 mM $CsMeSO_3$, 10 mM HEPES, 10 mM Na-phosphocreatine, 4 mM $MgCl_2$, 4 mM Na_2-ATP, 0.4 mM Na-GTP, 3 mM ascorbate, 0.03 mM Alexa 594 and 0.5 mM Fluo-4FF.

Synaptic stimulation (Fig. 3) was performed using short current pulses (0.1 ms, 30 μA) delivered with a glass pipette (~2–3 μm tip) filled with ACSF and 10 μM Alexa 594 to aid pipette placement. The pipette was positioned 10–20 μm from a GFP-labelled dendrite of interest. Each pulse produced an EPSP with amplitudes of 7.8 ± 2.7 mV ($n = 7$ stimulus positions from 3 cells, mean ± s.d., measured in parallel experiments), corresponding to the activation of ~30 synapses, a small subset of the synapses on typical CA1 pyramidal cells (total, ~10^4 synapses). The activated synapses are distributed throughout the dendritic tree, implying that multiple activated synapses are rarely found on a short stretch of dendrite[48,49]. We identified activated synapses that had undergone plasticity on the basis of spine enlargement. Image stacks containing a 30-μm-long stretch of dendrite were compared before and immediately after the synaptic LTP protocol (120 stimuli at 2 Hz, low extracellular Mg^{2+}). $\Delta F/F$ images (Fig. 3b) were generated after low-pass filtering and image alignment using cross-correlation analysis allowing for distortions. Spontaneous fluctuations in fluorescence intensity (that is, spine volume) in non-stimulated spines had a coefficient of variation of 0.21 ± 0.02. After the synaptic LTP protocol, spines that enlarged by more than three times the coefficient of variation of spontaneous fluctuations (ΔVol > 60%) were scored as synaptic LTP spines. Structural plasticity after synaptic stimulation was sparse, consistent with the expected activation of a small subset of synapses. Of the 114 imaged dendrites, 16 contained at least one enlarged spine (range 1–2 spines). For 14 of these 16 dendrites, only a single spine in the field of view enlarged after synaptic stimulation. In the two cases where multiple spines enlarged, the spine receiving the subthreshold protocol was less than 12 μm from one, but not from the other, enlarged spine.

Imaging and glutamate uncaging. Two-photon imaging and glutamate uncaging were performed using a custom-built microscope with two Ti:sapphire lasers (910 nm for imaging GFP and 720 nm for uncaging; MaiTai, Spectra Physics) controlled by ScanImage[50], as described[51]. In brief, the intensity of each beam was controlled independently by electro-optical modulators (Pockels cells, Conoptics). The polarization angle was set using a half-wave plate. The beams were combined with a polarizing beam-splitting cube (CVI Laser Optics) and passed through the same set of scan mirrors and objective (×60, 0.9 NA; Olympus). To aid alignment, two steering mirrors were used for each beam. The upstream position-steering mirror adjusted the position of the beam at the back focal plane of the objective. The downstream angle-steering mirror (ASM) was placed in a conjugate plane to the scan mirrors and back focal plane of the objective using a Keplerian telescope consisting of two long focal-length plano-convex lenses. Adjustment of the ASM changed the angle, but not the position, of the beam at the back focal plane of the objective, thus moving the beam in the sample plane. Coarse alignment was first performed to center the beams at the back focal plane of the objective. For fine alignment at the sample plane, 0.1 μm fluorescent beads were imaged simultaneously with both beams. The ASMs were adjusted until the images overlapped. The *x*, *y* and *z* resolutions (full-width at half-maximum) for the imaging beam (910 nm) were 0.53 μm, 0.59 μm and 1.89 μm, respectively. For the uncaging beam (720 nm), the *x*, *y* and *z* resolutions (full-width at half-maximum) were 0.50 μm, 0.56 μm and 1.66 μm, respectively.

Green and red fluorescence photons were separated using a dichroic mirror (565 nm) and bandpass filters (510/70, 635/90; Chroma). Photons were collected using photomultiplier tubes (Hamamatsu R3896 except for the epifluorescence green signal, which was collected using Hamamatsu H7422-40). Epi- and trans-fluorescence signals were collected and summed[52].

For glutamate uncaging, 2.5 mM MNI-caged-L-glutamate was added to the ACSF. Only spines well separated from both the dendrite and neighbouring spines were selected for experiments. The laser beam was parked at a manually determined uncaging location approximately 0.5 μm from the tip of the spine head in the direction away from the parent dendrite. The uncaging location was readjusted between uEPSC amplitude time points (5 test pulses per time point, 0.1 Hz; 5 min between time points), but not between test pulses (Figs 1b, d, f and 4d–f). The uncaging location was also repositioned after spine enlargement. To assess the accuracy of manual positioning, we compared uEPSC amplitudes across baseline time points for all spines (LTP, sub and nearby spines). uEPSC amplitudes were not significantly different between time points (repeated measures ANOVA, $P = 0.25$); similarly, trial-to-trial fluctuations were not different for test pulses within a time point and between time points (*t*-test, $P = 0.5$). Manual repositioning therefore did not affect the measurement of uEPSC amplitude nor did it contribute significantly to uEPSC amplitude fluctuations.

For test pulses, 45 mW laser power was delivered to the back focal aperture of the objective for 1 ms. During stimulus trains, we used 20-mW pulses lasting 4 ms for the LTP protocol and lasting 1 ms for the subthreshold protocol. For spike-timing-dependent LTP, all uncaging pulses were 1 ms in duration with 45 mW laser power at the back focal aperture of the objective. Approximately 20% of this laser power was transmitted through the objective. Initial spine volumes were indistinguishable across conditions (data not shown). The distances between spines tested for uEPSC changes were similar on average (Fig. 1: LTP protocol only, 3.1 ± 0.4 μm; subthreshold protocol only, 2.6 ± 0.6 μm; crosstalk, 3.5 ± 0.2 μm; ANOVA, $P > 0.7$. Figure 4: $\Delta t = 5$ ms, 2.2 ± 0.3 μm; $\Delta t = 35$ ms, 2.3 ± 0.4 μm; crosstalk, 2.3 ± 0.2 μm; ANOVA, $P > 0.95$). The depth in the slice was restricted to 25–50 μm.

Plasticity was induced using four protocols: depolarization to ~0 mV in perforated patch-clamp mode paired with 30 uncaging pulses at 0.5 Hz in 2 mM Ca^{2+}, 1 mM Mg^{2+} and 1 μM TTX (Fig. 1); 30 uncaging pulses at 0.5 Hz in 4 mM Ca^{2+}, 0 mM Mg^{2+} and 1 μM TTX (Fig. 2); 120 synaptic stimuli at 2 Hz in 4 mM Ca^{2+} and 0 mM Mg^{2+} at 33 °C (Fig. 3); and an uncaging pulse followed by 3 action potentials at 50 Hz, repeated 60 times at 2 Hz in 2 mM Ca^{2+} and 1 mM Mg^{2+} in perforated patch-clamp mode (Fig. 4). The time window (Δt) for spike-timing-dependent LTP was defined as the time between the uncaging pulse and the first action potential.

$[Ca^{2+}]$ imaging was performed as described[49]. Images were acquired every 64 ms in frame-scan mode. $[Ca^{2+}]$ transients were measured as the change in Ca^{2+}-sensitive green fluorescence (500 μM Fluo-4FF; ΔG) divided by the Ca^{2+}-insensitive red fluorescence (30 μM Alexa 594; R), normalized to $(G/R)_{max}$ measured in 10 mM Ca^{2+}.

Controls for inhibitor function. To test the efficacy of thapsigargin and ryanodine (Supplementary Fig 3), a CA1 cell in an acute hippocampal slice was filled with 500 μM Fluo 4FF and 30 μM Alexa 594. Caffeine (40 mM in ACSF) was pressure-applied for 2 sec from a pipette located ~20 μm from the soma of the filled cell[53]. Caffeine-induced $[Ca^{2+}]$ transients in the soma were measured before and 5 min after the application of 1 μM thapsigargin and 20 μM ryanodine.

To test protein synthesis inhibitor function (Supplementary Fig 4), cells in cultured rat hippocampal slices were transfected with destabilized EGFP[40]. Slices were incubated in ACSF at room temperature. Changes in green fluorescence intensity in the thick apical dendrite were monitored following application of DMSO (0.1 %), 25 μM anisomycin, 60 μM cycloheximide or 50 μM emetine.

Data analysis. Spine volumes were measured as the integrated green fluorescence after background subtraction, which is proportional to spine volume[45], normalized to the fluorescence intensity of the thick apical dendrite[25]. The origin of all time axes corresponds to the start of the uncaging protocols. Volume changes at nearby spines (Figs 1–4) were averaged across all neighbouring spines less than 4 μm from the LTP or sub spines. uEPSC changes at nearby spines (Figs 1 and 4) were from an individual neighbouring spine for each experiment. In the bar graphs, ΔVol and ΔuEPSC were normalized to the baseline and measured starting 15 min post stimulus until the end of the time course.

All data are presented as mean ± s.e.m. unless noted otherwise. *n* indicates the number of spines analysed. One experiment was performed per cell, except to map the spike-timing-dependent LTP time window for which up to three experiments per cell were performed sequentially. Each figure summarizes all experiments, except for Fig. 3 in which only experiments with a scored synaptic LTP spine in the field-of-view were analysed.

P values are from two-tailed *t*-tests unless noted otherwise. For all *t*-tests the null hypothesis stated that the mean was equal to zero, except for pharmacology experiments in which the null hypothesis stated that the means of control and drug conditions were the same.

46. Stoppini, L., Buchs, P. A. & Muller, D. A. A simple method for organotypic cultures of nervous tissue. *J. Neurosci. Methods* **37**, 173–182 (1991).
47. Yasuda, R. *et al.* Supersensitive Ras activation in dendrites and spines revealed by two-photon fluorescence lifetime imaging. *Nature Neurosci.* **9**, 283–291 (2006).
48. Oertner, T. G., Sabatini, B. S., Nimchinsky, E. A. & Svoboda, K. Facilitation at single synapses probed with optical quantal analysis. *Nature Neurosci.* **5**, 657–664 (2002).
49. Yasuda, R. *et al.* Imaging calcium concentration dynamics in small neuronal compartments. *Sci. STKE* **2004**, pl5 (2004).
50. Pologruto, T. A., Sabatini, B. L. & Svoboda, K. ScanImage: flexible software for operating laser-scanning microscopes. *Biomed. Eng. Online* **2**, 13 (2003).
51. Pologruto, T. A. *Imaging neural activity and [Ca^{2+}] with genetically encoded calcium indicators and two-photon excitation laser scanning microscopy.* PhD thesis, Harvard Univ. (2004).
52. Mainen, Z. F. *et al.* Two-photon imaging in living brain slices. *Methods* **18**, 231–239 (1999).
53. Garaschuk, O., Yaari, Y. & Konnerth, A. Release and sequestration of calcium by ryanodine-sensitive stores in rat hippocampal neurones. *J. Physiol.* **502**, 13–30 (1997).

Vol 450|20/27 December 2007|doi:10.1038/nature06396

ARTICLES

Structural basis for gate-DNA recognition and bending by type IIA topoisomerases

Ken C. Dong[1,2] & James M. Berger[2]

Type II topoisomerases disentangle DNA to facilitate chromosome segregation, and represent a major class of therapeutic targets. Although these enzymes have been studied extensively, a molecular understanding of DNA binding has been lacking. Here we present the structure of a complex between the DNA-binding and cleavage core of *Saccharomyces cerevisiae* Topo II (also known as Top2) and a gate-DNA segment. The structure reveals that the enzyme enforces a 150° DNA bend through a mechanism similar to that of remodelling proteins such as integration host factor. Large protein conformational changes accompany DNA deformation, creating a bipartite catalytic site that positions the DNA backbone near a reactive tyrosine and a coordinated magnesium ion. This configuration closely resembles the catalytic site of type IA topoisomerases, reinforcing an evolutionary link between these structurally and functionally distinct enzymes. Binding of DNA facilitates opening of an enzyme dimerization interface, providing visual evidence for a key step in DNA transport.

Topoisomerases resolve deleterious topological problems in chromosomes arising from transcription, repair and replication events[1]. These enzymes are divided into two classes, type I and type II, depending on whether they cleave one or two strands of DNA during their catalytic cycle. Type II topoisomerases, which are further divided into the IIA and IIB subclasses[2], promote the ATP-dependent passage of one DNA segment (the transport or T-segment) through a transient, double-stranded break in a second duplex (the gate or G-segment). G-segment cleavage is carried out by the nucleophilic attack of a pair of catalytic tyrosines on the phosphodiester backbone, forming a covalent enzyme•DNA intermediate. Small-molecule inhibitors can trap this cleavage complex, generating DNA lesions that promote cell death. Such compounds have been successfully exploited for developing an extensive number of antibacterial and anti-cancer treatments[3,4].

The reaction cycle for type IIA topoisomerases is believed to use a 'two-gate' mechanism for strand passage[5–8] (Supplementary Fig. 1). In this scheme, the T-segment enters through a protein gate controlled by the reversible dimerization of two ATPase domains on one side of the enzyme (the N-gate), and exits through a carboxy-terminal dimerization interface on the other (the C-gate). Passage of the T-segment through the cleaved G-segment is further coordinated by a third dimer interface (the DNA-gate) in the interior of the topoisomerase, formed by two winged-helix domains (WHDs) that contain the catalytic tyrosines responsible for DNA cleavage and separation[9]. A metal-binding TOPRIM domain, composed of a triad of conserved acidic residues[10], is believed to assist formation of the covalent protein•DNA intermediate by coordinating magnesium ions essential for DNA cleavage[11–13]. Interestingly, TOPRIM folds also form part of the catalytic centres of type IA topoisomerases and bacterial primases[10,14–17].

Despite extensive study, many fundamental aspects of type II topoisomerase function remain unresolved. For example, type IIA topoisomerases are known to associate preferentially with DNA crossovers and the apices of plectonemic supercoils[18–21], yet the molecular basis for this discrimination is not understood. The mechanisms by which G-segment binding can be uncoupled from cleavage in the absence of ATP, and the physical role of divalent metal ions in duplex scission[12,13], likewise remain unclear. Finally, a number of key strand passage intermediates predicted to exist from biochemical data have not been captured crystallographically. To address these and other outstanding issues, we determined the structure of the DNA-binding and cleavage core of a type IIA topoisomerase bound to prospective G-segment DNA.

Overview of the complex

The *S. cerevisiae* Topo II fragment used for this study (amino acids 408–1177) contains the metal-binding TOPRIM fold and primary DNA-binding domain (Fig. 1a), and can associate with and cleave DNA *in vitro*[22] (Supplementary Fig. 2). The structure was determined by X-ray crystallography to 3 ångström (Å) resolution in complex with a double-stranded DNA oligonucleotide containing complementary, four-base 5′ overhangs (Supplementary Fig. 3). The Topo II DNA-binding and cleavage core forms a 180 kDa dimer with overall dimensions of $100 \times 105 \times 80$ Å (Fig. 1b, c). The largest inter-subunit contacts occur through a predominantly polar and electrostatic set of interactions between the TOPRIM domain of one subunit and the DNA-binding domain of the other, burying ~2,150 Å^2 of solvent-accessible surface area per protomer (Supplementary Fig. 4a). A previously disordered linker region (amino acids 656–674) connecting the TOPRIM and WHD elements[9,23] is now clearly visible, forming a short three-helix bundle that makes extensive contacts with both DNA and the rest of the protein (Fig. 1b, c and Supplementary Fig. 4b).

Each 15 bp DNA oligonucleotide duplex binds a single Topo II protomer (Supplementary Fig. 5a). Contacts to the DNA are established within the confines of a large, positively charged groove that spans the width of the dimer and buries ~2,900 Å^2 of DNA surface area (Fig. 1d). The binding orientation of the DNA allows base pairing to occur between the four-base 5′ overhangs in the central portion of the DNA-binding site, generating a continuous, but doubly nicked, 34 bp duplex that crosses the enzyme dyad symmetry axis. The blunt duplex DNA ends emanating from each Topo II dimer stack against a second complex to form a compact 68 bp DNA circle in the crystal lattice (Supplementary Fig. 5b).

[1]Chemical Biology Graduate Program, Department of Chemistry, College of Chemistry, University of California, Berkeley, California 94720-3220, USA. [2]Division of Biochemistry and Molecular Biology, Department of Molecular and Cell Biology, QB3 Institute, Stanley Hall 3220, University of California at Berkeley, Berkeley, California 94720-3220, USA.

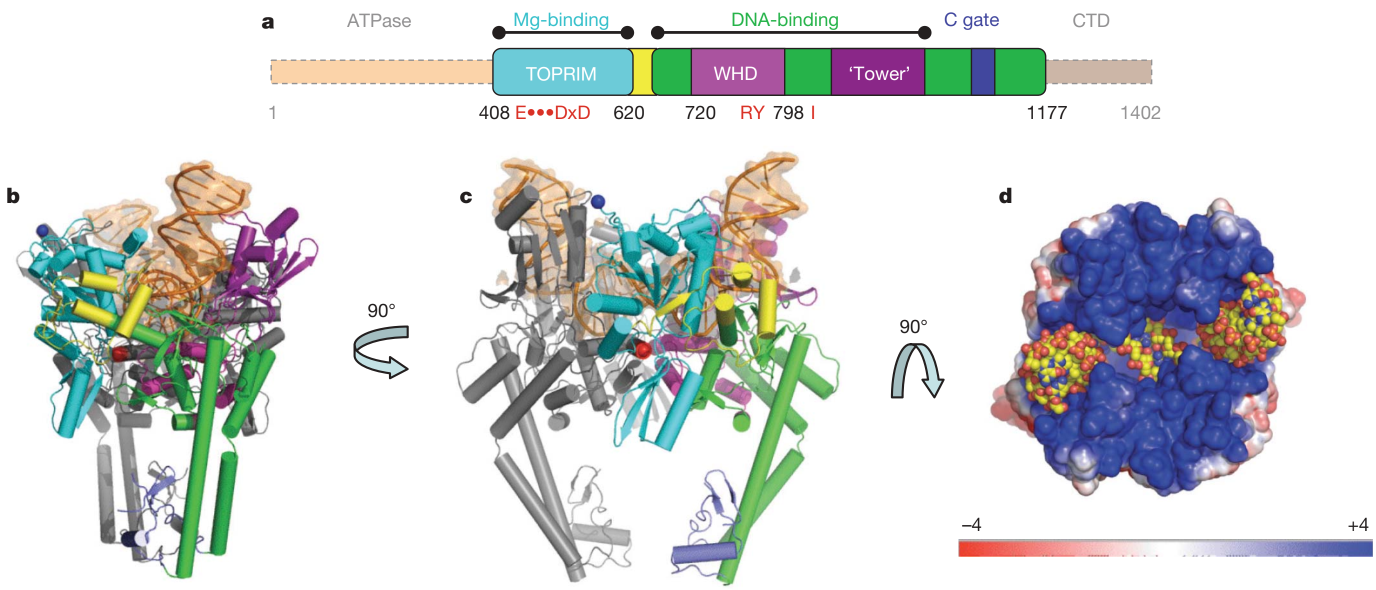


Figure 1 | Structure of the Topo II DNA-binding and cleavage core bound to DNA. **a**, Domain arrangement of *S. cerevisiae* Topo II: metal-binding TOPRIM (TOpoisomerase/PRIMase) fold, cyan; linker domain, yellow; WHD, magenta; tower domain, dark purple; dimerization (C-gate) domain, blue-grey; and auxiliary scaffolding, green. Regions not included in this study are faded. CTD, C-terminal domain. General locations of the metal-liganding residues (E and DxD), the catalytic tyrosine and accompanying arginine (RY), and DNA-bending isoleucine (I) are labelled. **b**, **c**, Cartoon representation of two orthogonal views of the complex. One protomer is coloured grey; the other is coloured as per panel **a**. DNA is shown in orange with a transparent surface. Spheres mark the locations of the catalytic tyrosines (red) and the amino termini (blue). **d**, Surface representation, coloured by electrostatic potential (red, negative; blue, positive) generated by the Adaptive Poisson-Boltzmann Solver (APBS)[38]. DNA is shown as spheres. Molecular figures were generated by PyMOL[39].

Topo II markedly reshapes DNA

One of the most distinctive features of the complex is the degree to which Topo II deforms bound DNA (Fig. 2a). As the DNA transitions into the catalytic interior, each subunit imparts a sharp 75° bend into its associated nucleic acid duplex. This distortion alters the trajectory of the duplex, creating a global bend of ~150° in the bound DNA segment. Between the two catalytic tyrosines in the cleavage centre, the DNA is essentially A form (Fig. 2a). In contrast, at the outermost edges of the G-segment-binding site, the duplex juts up and away from the protein, and is B form (Fig. 2b). Approximately 26 bp of the helix are within van der Waals distance of the protein—a value in agreement with footprinting data[24]. Consistent with the relatively low sequence specificity of Topo II, there are virtually no direct side-chain–base contacts (Supplementary Fig. 3b). The finding that the

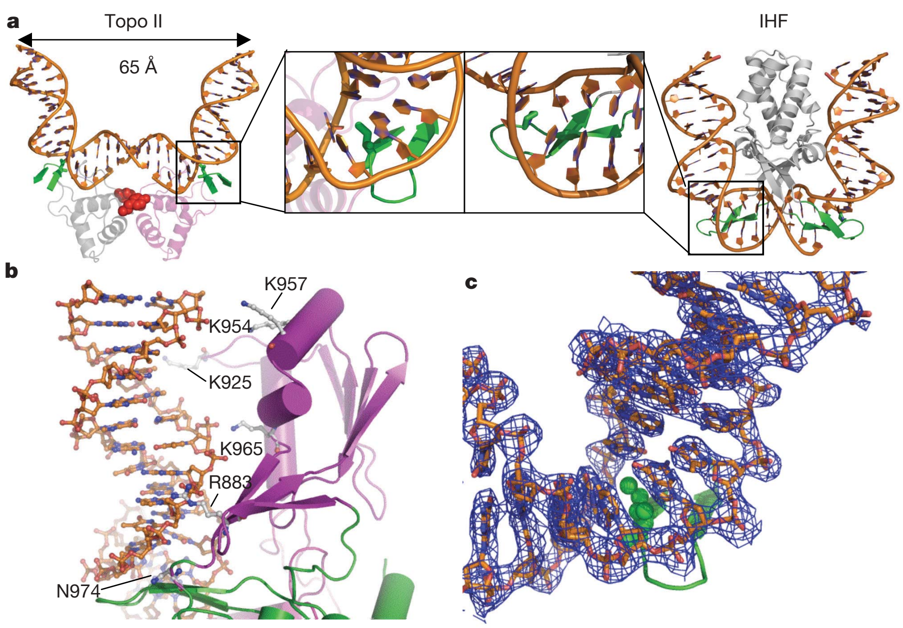


Figure 2 | Topo II severely bends DNA. **a**, Comparison of DNA bending caused by Topo II (left) and IHF[28] (right; Protein Data Bank (PDB) accession number, 1IHF). DNA-binding β-hairpins and intercalating aliphatic amino acids are highlighted in green and are shown in the inset. For Topo II, the WHDs and catalytic tyrosines are shown as magenta/grey ribbons and red spheres, respectively; the body of IHF is shown as a grey ribbon. **b**, DNA (orange) and tower domain (purple) interactions. Residues that contact DNA are shown as sticks. **c**, DNA-bending induced by Ile 833 (green spheres). Electron density from a simulated annealed omit map is shown contoured around the duplex at 1σ.

central four base pairs of the binding site are exposed and adopt an A-form conformation also agrees with hydroxyl radical footprinting data and the ability of ribonucleotides to enhance cleavage when placed within this region[25,26].

The interactions between Topo II and DNA reveal important roles for conserved structural elements in the G-segment-binding site. For example, the Topo II DNA-binding regions are demarcated by an $\alpha\beta$ fold (the 'tower' domain, residues 875–970) found in a number of DNA- and RNA-binding proteins (Class, Architecture, Topology, Homologous (CATH) classification: 3.30.1360). The precise function of this region had been unclear, but can now been seen to act as a buttressing element for the outer DNA ends (Fig. 2b), with basic residues on one face of the fold making polar and electrostatic contacts with the phosphate backbone of DNA.

The physical basis for the sharp DNA bend also is evident. Resident on the surface of each protomer's DNA-binding groove is a β-hairpin (amino acids 830–838) that contains a surface-exposed isoleucine (Ile 833) (Figs 2a, c). This residue, which is invariant across all but one type IIA topoisomerase homologue (African Swine Fever Virus, where it is a proline), intercalates between base pairs at the +8 and +9 positions of the bound DNA, significantly widening the minor groove by increasing the rise and roll of the duplex and decreasing local twist (Supplementary Fig. 6). The bent conformation is further stabilized by a deep, positively charged groove formed near the WHD. This feature helps to channel the DNA backbone towards the active site through side-chain–phosphate interactions, most notably an invariant histidine (His 736) and lysine (Lys 700).

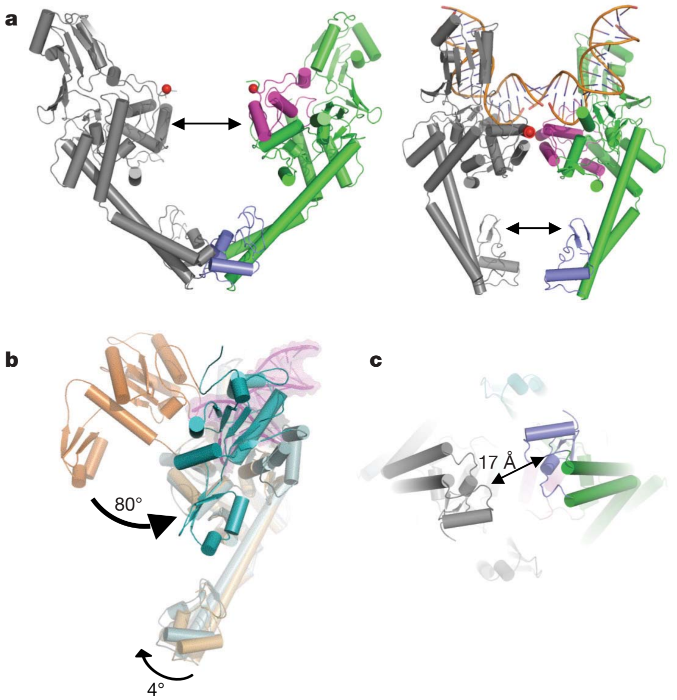


Figure 3 | Topo II conformational changes accompany DNA binding. **a**, Comparison of the WHD–WHD (DNA-gate) and CTD–CTD (C-gate) dimer interactions between the crystal structures of the DNA-binding domain of apo Topo II (left; PDB accession number, 1BGW) and Topo II bound to DNA (right). For clarity, the TOPRIM domains have been omitted from the figure; other regions are coloured as in Fig. 1. Arrows indicate separation points between inter-subunit contacts. **b**, Superposition of the DNA-binding regions (residues 700–1200, grey cylinders) of apo and DNA-bound Topo II shows a large (~80°) *en bloc* rotation of the metal-binding TOPRIM fold (dark orange, apo; blue, DNA complex). A modest 4° swivel of the C-terminal dimerization domain is also seen. DNA is coloured magenta. **c**, C-gate separation. The view is looking up from the bottom of the molecule shown in the right-hand-side of panel **a**.

Mutation of this latter residue to alanine is known to inactivate Topo II (ref. 27).

The U turn caused by Topo II is reminiscent of the deformations induced by the DNA-remodelling protein integration host factor (IHF)[28]. Although the overall structures of IHF and Topo II are completely unrelated, both proteins use a pair of β-hairpin 'arms' with a conserved aliphatic amino acid (isoleucine or proline) that intercalates into the DNA minor groove. The combined action of two hairpins from a protein dimer, along with stabilizing contacts to the flanking arms outside of the bent region, allows IHF to stabilize a global DNA bend similar in scope to that seen for Topo II (~180° versus 150°) (Fig. 2a). An analysis of the bound oligonucleotides using 3DNA[29], which quantitatively assesses DNA structure and stereochemistry, reveals that IHF and Topo II exert similar effects on the rise, roll and twist of the DNA region around the bend (Supplementary Fig. 6).

Extensive enzyme rearrangements accompany DNA binding

In addition to deforming DNA, Topo II undergoes substantial conformational changes on binding substrate. Structures of apo *S. cerevisiae* Topo II have shown that the two dimer-related WHDs and their resident active site tyrosines can separate by 25–30 Å (Fig. 3a)[9,23]. In the presence of DNA, however, these elements move towards one another, forming a nearly contiguous interaction surface across both protomers. This conformation is similar to that seen for the isolated GyrA subunit of DNA gyrase[30], a bacterial type IIA topoisomerase, except that in Topo II there are only peripheral WHD–WHD interactions. Instead, the TOPRIM domains of each Topo II protomer rotate by ~80° to dock against the tower and WHD of their partner subunit (Fig. 3b), thereby physically blocking more substantive WHD–WHD contacts. The total area buried in the TOPRIM–WHD–tower interaction is also significantly larger than that in the WHD–WHD contacts seen in the GyrA dimer (~2,900 Å^2 versus 1,300 Å^2 per protomer). Several residues conserved in eukaryotes on the tower domain (Asp 932, Glu 935 and Lys 913) form electrostatic contacts with conserved residues on the metal-binding fold (Arg 422, Asn 433, Glu 431 and Arg 466), indicating that this interface is functionally significant (Supplementary Fig. 4).

A conformational change accompanying DNA binding also takes place in the C-terminal dimerization interface of the protein (Fig. 3a). This region is linked to the main body of Topo II by an extended, α-helical coiled-coil that flexes to accommodate movements in the

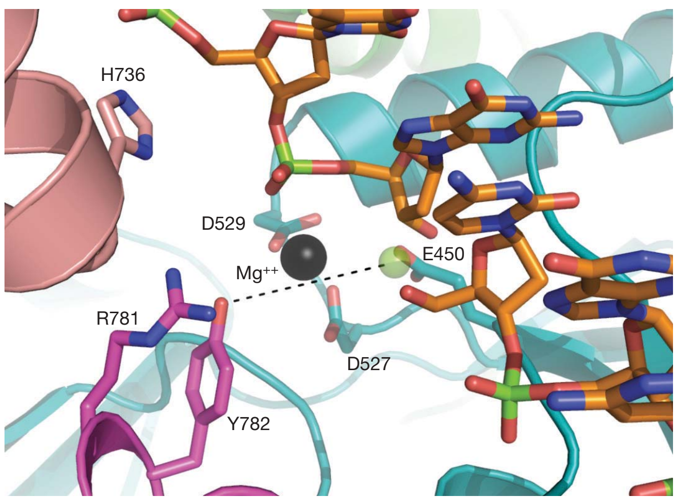


Figure 4 | Close up of the Topo II active site. The TOPRIM fold and WHD of one protomer are coloured cyan and pink, respectively, whereas the WHD of the partner subunit is coloured magenta. Key residues are shown as sticks and are labelled. A green sphere marks the midpoint of the nick between the +1 and −1 position of double-stranded DNA, and corresponds to the approximate position of the scissile phosphodiester linkage.

globular DNA-binding domains that drive G-segment opening[23,30]. In previous type IIA topoisomerase structures, the C-gate interface can be seen to form a well-packed inter-subunit contact point[9,23,30]. In the structure of the complex, however, G-segment binding is accompanied by separation of the interface, opening a ~17 Å gap between partner subunits (Fig. 3c).

The reorientation of metal-binding folds in the DNA-binding region leads to the formation of a bipartite active site between Topo II subunits. This conformation positions a triad of conserved acidic residues (Glu 450, Asp 527 and Asp 529) in the TOPRIM domain close to the catalytic tyrosine, Tyr 782 (Fig. 4). Inspection of electron density maps revealed a large single peak of positive difference density associated with these acidic residues and the DNA phosphodiester backbone; on the basis of the peak height for this feature ($>7\sigma$) and its proximity to several liganding groups (<2.5 Å), this density was modelled as a Mg^{2+} ion. Interestingly, Tyr 782 and its invariant partner amino acid, Arg 781, lie just outside (7 Å) of the hydrogen bonding or van der Waals contact distance of DNA backbone, indicating that the active site has not yet attained a cleavage-competent configuration (Fig. 4).

Notably, the reorganization of the TOPRIM and WHD helps account for genetic and biochemical data showing that two groups of drug-resistance mutations segregated spatially in apo enzyme models actually form a composite interaction locus in the presence of DNA (Supplementary Fig. 7). The structural changes also give rise to an active site configuration for Topo II that is highly similar to the single-strand-cleaving type IA topoisomerases (Supplementary Fig 8). This congruence helps explain several equivalent catalytic properties between the two architecturally and functionally distinct classes of enzyme, such as the ability to form 5′ phosphotyrosine linkages with nucleic acid segments[31] and a requirement for metal ions for full activity[12,13,32]. This observation indicates that type IA and type IIA topoisomerases evolved from a common ancestor and probably utilize similar mechanisms for cleaving DNA.

Mechanistic implications

Together, the types of molecular contacts and inter-domain conformations we observe for this DNA-bound state of Topo II substantiate a number of previous biochemical findings and establish the structure as a meaningful intermediate for considering the type IIA topoisomerase mechanism. From this vantage, the structure provides a window into understanding a number of important functional issues.

Control of DNA cleavage. Although type IIA topoisomerases can cleave DNA to some extent in the absence of ATP, nucleotide binding greatly stimulates the reaction[33,34]. This behaviour indicates that, following binding of G-segment DNA, Topo II can enter into an equilibrium between cleaved and uncleaved states, with the uncleaved state favoured in the absence of ATP and the cleaved state promoted by the addition of nucleotide. One physical explanation for the uncoupling of binding and cleavage is immediately suggested by the structure of the DNA-bound complex. Association with a G-segment occurs concomitantly with formation of a bipartite active site consisting of the WHDs and the TOPRIM metal-binding region (Figs 1 and 4). This event positions the catalytic tyrosines adjacent to the scissile phosphodiester linkage, but at a distance insufficient to support nucleophilic attack. Instead, a small structural change—either compression of the DNA into the active site or a subtle reorientation of catalytic residues—appears necessary to position the tyrosine close enough to DNA to initiate cleavage. We currently cannot rule out the possibility that the observed configuration of the tyrosine results from the absence of a scissile phosphate in the cleavage position of our bound oligonucleotide substrate. Nevertheless, given the myriad of features consistent with available biochemical data, the active site seems likely to represent a 'pre-catalytic', but inactive, conformation that is poised to interconvert into a cleavage-competent state.

Mechanics of DNA transport. Although studies on both Topo II and DNA gyrase have suggested that the C-terminal dimerization domains of the primary DNA-binding region can separate to expel a T-segment following its transport through the G-segment[5,7,8], direct observation of this state had been lacking. The structure of the complex now shows that the C-terminal dimer interface can indeed open (Fig. 3), providing concrete physical evidence for this key intermediate. The structure also indicates that a T segment does not have to be present for the C-gate to open. Although the distance between the C-gate interface is slightly smaller than the diameter of a DNA duplex (Fig. 3c), the coiled-coil elements that link the dimerization domains to the rest of the protein are known to flex significantly[23,30], suggesting that this gap can be widened further to promote T-segment release (Supplementary Fig. 9 and Supplementary Movie 1).

Binding site preference and global control of DNA topology. The ability of type IIA topoisomerases to deform DNA has been the subject of much debate. Electron microscopy studies have shown that these enzymes associate preferentially with regions of significant DNA curvature, including the apices of plectonemic supercoils and DNA crossovers[18,19,21]. Interpretations of these and other lines of biochemical data have been somewhat conflicted, however, with some groups reporting significant bending of DNA by type IIA topoisomerases[21] and others finding no evidence for such distortions[35]. Our structure unequivocally resolves this question, showing that Topo II can bend DNA by more than 150° (Fig. 2).

DNA bending has important biological ramifications for the mechanism of type IIA topoisomerases, particularly for a puzzling phenomenon known as topology simplification[36]. This process relates to the ability of type IIA topoisomerases to actively reduce the concentrations of supercoiled topoisomers and catenated DNAs to levels below those produced at thermodynamic equilibrium, and is thought to help ensure that DNA segments are fully disentangled before chromosome partitioning. Although numerous models have been advanced to account for simplification behaviour, our structure strongly supports those that invoke G-segment bending as a means to promote decatenation and relaxation[21,37]. The structure further explains why type IIA topoisomerases tend to associate with highly curved DNA crossovers and loops, and suggests that preferred binding sites observed *in vivo* probably arise from a high intrinsic degree of deformability in the nucleic acid duplex.

Conclusions

The structural studies presented here reveal new insights into several aspects of type IIA topoisomerases, including how G-segment recognition is linked to but uncoupled from cleavage, and why bent DNA regions serve as favoured binding sites for the enzymes. The structure also provides visual evidence for a key conformational intermediate essential for DNA transport and establishes a firm evolutionary connection between type II and IA topoisomerases. This work should serve as a useful platform for future efforts to understand the chemical basis of DNA cleavage, and for drug development endeavours aimed at understanding and improving anti-topoisomerase therapeutics.

METHODS SUMMARY

The DNA-binding and cleavage domain was expressed in yeast and subsequently purified using a nickel-affinity, ion-exchange and size-exclusion chromatography. DNA–protein complexes were formed by dialysis, and crystals were grown by hanging-drop vapour diffusion. Diffraction data were collected from cryo-preserved crystals on BL8.3.1 (Advanced Light Source, Lawrence Berkeley National Laboratories). Molecular replacement of individual topoisomerase domains was used to obtain starting phases for model building; clear F_o–F_c difference density was evident for DNA at this point (Supplementary Fig. 4a). The final model, which contains all DNA nucleotides and amino acids 419–1177 of the Topo II fragment, was refined to an R_{work} of 22.9% and R_{free} of 27.7% with good stereochemistry (Supplementary Table 1). A more detailed explanation of our procedures can be found in the Methods section.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 20 July; accepted 15 October 2007.

1. Wang, J. C. Cellular roles of DNA topoisomerases: a molecular perspective. *Nature Rev. Mol. Cell Biol.* **3,** 430–440 (2002).
2. Gadelle, D., Filee, J., Buhler, C. & Forterre, P. Phylogenomics of type II DNA topoisomerases. *Bioessays* **25,** 232–242 (2003).
3. Hande, K. R. Clinical applications of anticancer drugs targeted to topoisomerase II. *Biochim. Biophys. Acta* **1400,** 173–184 (1998).
4. Hooper, D. C. & Rubinstein, E. *Quinolone antimicrobial agents* xiv (ASM, Washington DC, 2003).
5. Roca, J., Berger, J. M., Harrison, S. C. & Wang, J. C. DNA transport by a type II topoisomerase: direct evidence for a two-gate mechanism. *Proc. Natl Acad. Sci. USA* **93,** 4057–4062 (1996).
6. Roca, J. & Wang, J. C. The capture of a DNA double helix by an ATP-dependent protein clamp: a key step in DNA transport by type II DNA topoisomerases. *Cell* **71,** 833–840 (1992).
7. Roca, J. & Wang, J. C. DNA transport by a type II DNA topoisomerase: evidence in favor of a two-gate mechanism. *Cell* **77,** 609–616 (1994).
8. Williams, N. L. & Maxwell, A. Probing the two-gate mechanism of DNA gyrase using cysteine cross-linking. *Biochemistry* **38,** 13502–13511 (1999).
9. Berger, J. M., Gamblin, S. J., Harrison, S. C. & Wang, J. C. Structure and mechanism of DNA topoisomerase II. *Nature* **379,** 225–232 (1996).
10. Aravind, L., Leipe, D. D. & Koonin, E. V. Toprim—a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. *Nucleic Acids Res.* **26,** 4205–4213 (1998).
11. Noble, C. G. & Maxwell, A. The role of GyrB in the DNA cleavage–religation reaction of DNA gyrase: a proposed two metal-ion mechanism. *J. Mol. Biol.* **318,** 361–371 (2002).
12. Osheroff, N. Role of the divalent cation in topoisomerase II mediated reactions. *Biochemistry* **26,** 6402–6406 (1987).
13. Goto, T., Laipis, P. & Wang, J. C. The purification and characterization of DNA topoisomerases I and II of the yeast *Saccharomyces cerevisiae. J. Biol. Chem.* **259,** 10422–10429 (1984).
14. Lima, C. D., Wang, J. C. & Mondragon, A. Three-dimensional structure of the 67K N-terminal fragment of *E. coli* DNA topoisomerase I. *Nature* **367,** 138–146 (1994).
15. Berger, J. M., Fass, D., Wang, J. C. & Harrison, S. C. Structural similarities between topoisomerases that cleave one or both DNA strands. *Proc. Natl Acad. Sci. USA* **95,** 7876–7881 (1998).
16. Podobnik, M., McInerney, P., O'Donnell, M. & Kuriyan, J. A. TOPRIM domain in the crystal structure of the catalytic core of *Escherichia coli* primase confirms a structural link to DNA topoisomerases. *J. Mol. Biol.* **300,** 353–362 (2000).
17. Keck, J. L., Roche, D. D., Lynch, A. S. & Berger, J. M. Structure of the RNA polymerase domain of *E. coli* primase. *Science* **287,** 2482–2486 (2000).
18. Howard, M. T., Lee, M. P., Hsieh, T. S. & Griffith, J. D. *Drosophila* topoisomerase II–DNA interactions are affected by DNA structure. *J. Mol. Biol.* **217,** 53–62 (1991).
19. Zechiedrich, E. L. & Osheroff, N. Eukaryotic topoisomerases recognize nucleic acid topology by preferentially interacting with DNA crossovers. *EMBO J.* **9,** 4555–4562 (1990).
20. Roca, J., Berger, J. M. & Wang, J. C. On the simultaneous binding of eukaryotic DNA topoisomerase II to a pair of double-stranded DNA helices. *J. Biol. Chem.* **268,** 14250–14255 (1993).
21. Vologodskii, A. V. *et al.* Mechanism of topology simplification by type II DNA topoisomerases. *Proc. Natl Acad. Sci. USA* **98,** 3045–3049 (2001).
22. Berger, J. M. *Structural determination of a 92 kDa fragment of yeast topoisomerase II by X-ray crystallography at 2.7 Å resolution.* PhD thesis, Harvard University (1995).
23. Fass, D., Bogden, C. E. & Berger, J. M. Quaternary changes in topoisomerase II may direct orthogonal movement of two DNA strands. *Nature Struct. Biol.* **6,** 322–326 (1999).
24. Thomsen, B. *et al.* Characterization of the interaction between topoisomerase II and DNA by transcriptional footprinting. *J. Mol. Biol.* **215,** 237–244 (1990).
25. Orphanides, G. & Maxwell, A. Evidence for a conformational change in the DNA gyrase–DNA complex from hydroxyl radical footprinting. *Nucleic Acids Res.* **22,** 1567–1575 (1994).
26. Wang, Y., Thyssen, A., Westergaard, O. & Andersen, A. H. Position-specific effect of ribonucleotides on the cleavage activity of human topoisomerase II. *Nucleic Acids Res.* **28,** 4815–4821 (2000).
27. Liu, Q. & Wang, J. C. Identification of active site residues in the "GyrA" half of yeast DNA topoisomerase II. *J. Biol. Chem.* **273,** 20252–20260 (1998).
28. Rice, P. A., Yang, S., Mizuuchi, K. & Nash, H. A. Crystal structure of an IHF–DNA complex: a protein-induced DNA U-turn. *Cell* **87,** 1295–1306 (1996).
29. Lu, X. J. & Olson, W. K. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **31,** 5108–5121 (2003).
30. Morais Cabral, J. H. *et al.* Crystal structure of the breakage–reunion domain of DNA gyrase. *Nature* **388,** 903–906 (1997).
31. Tse, Y. C., Kirkegaard, K. & Wang, J. C. Covalent bonds between protein and DNA. Formation of phosphotyrosine linkage between certain DNA topoisomerases and DNA. *J. Biol. Chem.* **255,** 5560–5565 (1980).
32. Domanico, P. L. & Tse-Dinh, Y. C. Mechanistic studies on *E. coli* DNA topoisomerase I: divalent ion effects. *J. Inorg. Biochem.* **42,** 87–96 (1991).
33. Sugino, A., Higgins, N. P., Brown, P. O., Peebles, C. L. & Cozzarelli, N. R. Energy coupling in DNA gyrase and the mechanism of action of novobiocin. *Proc. Natl Acad. Sci. USA* **75,** 4838–4842 (1978).
34. Osheroff, N. Eukaryotic topoisomerase II. Characterization of enzyme turnover. *J. Biol. Chem.* **261,** 9944–9950 (1986).
35. Trigueros, S., Salceda, J., Bermudez, I., Fernandez, X. & Roca, J. Asymmetric removal of supercoils suggests how topoisomerase II simplifies DNA topology. *J. Mol. Biol.* **335,** 723–731 (2004).
36. Rybenkov, V. V., Ullsperger, C., Vologodskii, A. V. & Cozzarelli, N. R. Simplification of DNA topology below equilibrium values by type II topoisomerases. *Science* **277,** 690–693 (1997).
37. Buck, G. R. & Zechiedrich, E. L. DNA disentangling by type-2 topoisomerases. *J. Mol. Biol.* **340,** 933–939 (2004).
38. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98,** 10037–10041 (2001).
39. DeLano, W. The PyMOL molecular graphics system. ⟨http://www.pymol.org⟩ (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors are grateful to D. Herschlag for advice on the choice of DNA substrate, and to members of the Berger laboratory for discussions. This work was supported by a NIH Training Grant position to K.C.D. and by the NCI.

Author Contributions K.C.D. and J.M.B. designed the experimental plan, analysed the data, and wrote the paper. K.C.D. performed the research.

Author Information Coordinates have been deposited in the RSCB PDB under the accession number 2RGR. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.M.B. (jmberger@berkeley.edu).

METHODS

Expression of the Topo II construct. An open reading frame encoding the DNA-binding and cleavage core of *S. cerevisiae* Topo II (topo IIΔNC, missing the N-terminal ATPase domain and the C-terminal phosphodomain, corresponding to amino acids 408–1177) was amplified by PCR using primers to add an N-terminal hexa-histidine tag and an accompanying Tobacco Etch Virus (TEV) protease site. This DNA was cloned into a pGAL vector used for expressing full-length Topo II (plasmid name pG1T2)[40], and the resulting plasmid was transformed in BCY123 yeast cells. After growth of cells at 30 °C in CSM URA^- drop-out media (Sunrise Science) to mid-log phase ($A_{600} \cong 1.0$), galactose ($20\,g\,l^{-1}$) was added and protein expression was induced for 6 h. Cells were harvested by centrifugation and were frozen in liquid nitrogen for storage.

Purification of Topo IIΔNC. For lysis, cells were crushed under liquid nitrogen using a mortar and pestle, the powder was resuspended in buffer A (20 mM Tris 8.5 and 10% glycerol) plus 20 mM imidazole and 300 mM KCl, and the lysate was clarified by centrifugation. Clarified extract was loaded onto a pair of tandemly coupled nickel affinity and cation-exchange columns (nickel-chelating Sepharose resin and Hi-Trap S, GE-Amersham Biosciences), which were then washed with buffer A plus 20 mM imidazole and 300 mM KCl to remove unbound proteins. Topo IIΔNC was eluted from the nickel column directly onto the ion-exchange column using buffer A plus 100 mM KCl and 200 mM imidazole, after which the columns were uncoupled, and the protein eluted from the ion-exchange column in buffer A using a 100–500 mM KCl gradient. Purified enzyme fractions were pooled, mixed with His_6-tagged TEV-protease[41] at a 10:1 molar ratio of protein:protease, and incubated at 4 °C for 18 h while concentrating in an Amicon Centriprep-30 concentrator to remove the His tag. TEV protease and uncleaved Topo II were removed by passing the concentrated solution over a second nickel-affinity column using buffer A with 500 mM KCl and 20 mM imidazole. The flow-through from this step (which contained TEV-cleaved Topo IIΔNC) was purified further over an S300 sizing column (GE Amersham) in 500 mM KCl, 50 mM Tris 7.5, 5 mM $MgCl_2$ and 10% glycerol, before concentrating and storing at 4 °C. The final concentration of Topo IIΔNC varied, but was always $>10\,mg\,ml^{-1}$. Similar to ATP-free Topo II, the purified Topo IIΔNC fragment protein exhibits a low but demonstrable DNA-cleavage activity *in vitro* (Supplementary Fig. 2)[22].

Purification of DNA substrate. Oligonucleotides were ordered from Operon Biosciences, resuspended in 100% formamide, and run on a 8.75 M urea, 10% polyacrylamide gel for 1 h at 350 V. DNA bands were identified by ultraviolet-shadowing, cut out, and the separated oligonucleotides eluted by crushing and soaking in 10 mM magnesium acetate, 500 mM ammonium acetate and 1 mM EDTA overnight at room temperature (20–25 °C). Gel-purified samples were further filtered using Steriflips (Millipore), and were loaded onto a Sep-Pak column (Waters) equilibrated in 100 mM triethylammonium bicarbonate (TEAB, Sigma), before washing the DNA with 100 mM TEAB and eluting with 30% acetonitrile in 25 mM TEAB. Column-purified DNAs were lyophilized, and were resuspended in 200 mM KCl, 10 mM Tris 7.5 and 10 mM $MgCl_2$ to a final concentration of 5.5 mM. For annealing, complementary oligos were mixed, heated to 95 °C and cooled slowly overnight to room temperature in a water bath.

Crystallization of the Topo II–DNA complex. For crystallization, $7.5\,mg\,ml^{-1}$ of purified Topo IIΔNC was first mixed with 1 mM freshly prepared sodium orthovanadate (Sigma) and annealed DNA duplexes at a 1:1.3 monomer:DNA ratio in a starting buffer of 50 mM Tris 7.5, 500 mM KCl and 2 mM $MgCl_2$. Sodium orthovanadate was added in an attempt to capture a transition state complex[42], but examination of electron density maps revealed no evidence for vanadate at the 3′–5′ junctions between oligonucleotide ends. Vandate-treated Topo IIΔNC•DNA solutions were dialyzed step-wise against 50 mM Tris 7.5, 2 mM $MgCl_2$ and decreasing amounts of KCl (400–200 mM) using 1 kDa cut-off micro DispoDIALYZERs (Harvard) to a final solution of 200 mM KCl, 50 mM Tris (pH 7.5) and 2 mM $MgCl_2$ over the course of 18 h at 4 °C. A high-molecular-weight osmolyte (2.5% (w/v) PEG 5000 monomethyl ether) was also included in the dialysis buffer to minimize sample dilution. For crystallization, the dialyzed sample was mixed with well solution (12–20% PEG 1000, 100–250 mM $MgCl_2$ and 100 mM sodium cacodylate, pH 7.0) at a 1:1 ratio, and was subjected to hanging-drop vapour diffusion at 4 °C. Over the course of a few days, rod-shaped crystals belonging to the space group *I*222 grew in heavy precipitate; despite extensive attempts to optimize growth, only 5–10 screenable crystals of variable quality could be obtained from 10–15 mg of starting material out of each 6 litres of yeast preparation. Crystals were harvested by transferring into 25% glycerol plus well solution before looping and flash-freezing in liquid nitrogen. For the native data set used in this study, a 2 h soak in harvest solution plus 1 mM $TmCl_3$ also was performed in an attempt to find a heavy atom derivative for experimental phasing. No anomalous difference density for thulium was observed, but the data were of sufficient quality to use for molecular replacement and refinement.

Structure determination of the Topo II–DNA complex. Diffraction data were collected on beamline 8.3.1 at the Advanced Light Source of Lawrence Berkeley National Laboratory[43] at 1.3776 Å and 100 K using an ADSC Q210 CCD detector, and were processed using HKL2000 (ref. 44; Supplementary Table 1). The structure was solved by molecular replacement with PHASER[45], as implemented under the CCP4i program suite[46], using the individual structures of the core DNA-binding domain of *S. cerevisiae* Topo II (PDB accession number, 1BGW; amino acids 683–1009 and 1151–1178)[9] and the conserved secondary structural elements of the TOPRIM fold of *E. coli* topoisomerase III (PDB accession number, 1I7D; amino acids 2–39 and 82–141)[47] as search models. Interestingly, searches using the isolated Topo II TOPRIM domain did not work initially for molecular replacement. Inspection of the refined structure revealed that residues 613–620, which had comprised an α-helix in the Topo II search model, transitioned to form a β-sheet in the Topo IIΔNC•DNA complex; because this region was absent in the Topo III TOPRIM fold, the Topo II TOPRIM may have constituted a overly divergent starting model.

Although interpretable electron density for DNA was evident immediately after molecular replacement of the isolated Topo II DNA-binding domain (Supplementary Fig. 4a), we did not attempt to model the oligonucleotide until after all protein regions had been either placed or manually built. Consistent with our modelling, attempts to use short B-form duplexes (10 bp) for molecular replacement (by means of PHASER) positioned the DNA in the same location as was evident visually from F_o–F_c difference density maps. The structure of the Topo IIΔNC•DNA complex was built using Coot[48] and refined using PHENIX[49] to a final R_{work} of 22.9% and R_{free} of 27.7%. No residues fall in disallowed regions of Ramachandran space (90% preferred, 100% allowed), as reported by Molprobity[50].

40. Worland, S. T. & Wang, J. C. Inducible overexpression, purification, and active site mapping of DNA topoisomerase II from the yeast *Saccharomyces cerevisiae*. *J. Biol. Chem.* **264**, 4412–4416 (1989).
41. Kapust, R. B. & Waugh, D. S. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* **8**, 1668–1674 (1999).
42. Davies, D. R. & Hol, W. G. The power of vanadate in crystallographic investigations of phosphoryl transfer enzymes. *FEBS Lett.* **577**, 315–321 (2004).
43. MacDowell, A. A. *et al.* Suite of three protein crystallography beamlines with single superconducting bend magnet as the source. *J. Synchrotron Radiat.* **11**, 447–455 (2004).
44. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
45. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. Likelihood-enhanced fast translation functions. *Acta Crystallogr. D* **61**, 458–464 (2005).
46. Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D* **59**, 1131–1137 (2003).
47. Changela, A., DiGate, R. J. & Mondragon, A. Crystal structure of a complex of a type IA DNA topoisomerase with a single-stranded DNA molecule. *Nature* **411**, 1077–1081 (2001).
48. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
49. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
50. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).

LETTERS

Late formation and prolonged differentiation of the Moon inferred from W isotopes in lunar metals

M. Touboul[1], T. Kleine[1], B. Bourdon[1], H. Palme[2] & R. Wieler[1]

The Moon is thought to have formed from debris ejected by a giant impact with the early 'proto'-Earth[1] and, as a result of the high energies involved, the Moon would have melted to form a magma ocean. The timescales for formation and solidification of the Moon can be quantified by using ^{182}Hf–^{182}W and ^{146}Sm–^{142}Nd chronometry[2–4], but these methods have yielded contradicting results. In earlier studies[3,5–7], ^{182}W anomalies in lunar rocks were attributed to decay of ^{182}Hf within the lunar mantle and were used to infer that the Moon solidified within the first ~60 million years of the Solar System. However, the dominant ^{182}W component in most lunar rocks reflects cosmogenic production mainly by neutron capture of ^{181}Ta during cosmic-ray exposure of the lunar surface[3,7], compromising a reliable interpretation in terms of ^{182}Hf–^{182}W chronometry. Here we present tungsten isotope data for lunar metals that do not contain any measurable Ta-derived ^{182}W. All metals have identical ^{182}W/^{184}W ratios, indicating that the lunar magma ocean did not crystallize within the first ~60 Myr of the Solar System, which is no longer inconsistent with Sm–Nd chronometry[8–11]. Our new data reveal that the lunar and terrestrial mantles have identical ^{182}W/^{184}W. This, in conjunction with ^{147}Sm–^{143}Nd ages for the oldest lunar rocks[8–11], constrains the age of the Moon and Earth to 62^{+90}_{-10} Myr after formation of the Solar System. The identical ^{182}W/^{184}W ratios of the lunar and terrestrial mantles require either that the Moon is derived mainly from terrestrial material or that tungsten isotopes in the Moon and Earth's mantle equilibrated in the aftermath of the giant impact, as has been proposed to account for identical oxygen isotope compositions of the Earth and Moon[12].

We obtained tungsten isotope data for metals from two KREEP-rich samples (KREEP stands for enrichment in potassium (K), rare earth elements (REE) and phosphorus (P)), four low-Ti and five high-Ti mare basalts (Fig. 1 and Table 1). We processed fourfold to fivefold more material than an earlier study[3] and monitored the purity of our metal separates by determining their Hf/W ratios. These indicate that for the analyses reported here any possible contamination from silicate and oxide grains has no measurable effect on ^{182}W/^{184}W. Most of the samples investigated here had relatively short exposure times and required corrections of <0.1 ε units ($\varepsilon = 0.01\%$) for burnout of tungsten isotopes[13,14]; only for samples 15556 and 70017 (exposure ages ~220 and ~500 Myr) were corrections larger (~0.4 and ~0.2 ε units). Details of the corrections are given in Supplementary Information. All samples analysed here have identical ^{182}W/^{184}W ratios within ±0.32 ε units (2 σ) and agree with previously reported data for metals from KREEP-rich samples[3]. Combined, these data average at $\varepsilon^{182}\mathrm{W} = 0.09 \pm 0.10$ (2 s.e.m.), $n = 15$; ε^{182}W is defined in Table 1).

In contrast to earlier studies[3,5,6], we do not find ^{182}W/^{184}W variations within the lunar mantle. Lee *et al.*[5] reported $\varepsilon^{182}\mathrm{W} \approx 1.4$ for low-Ti mare basalt 15555; however, the exposure age of this sample, combined with its Sm isotopic composition and Ta/W ratio, indicates that this anomaly might be due entirely to cosmogenic ^{182}W. Kleine *et al.*[3] reported elevated $\varepsilon^{182}\mathrm{W} \approx 2$ for a magnetic separate from high-Ti mare basalt 79155 but we determined Hf/W = 7.5 for an aliquot from the same magnetic separate, most probably indicating the presence of some ilmenite and hence cosmogenic ^{182}W in this separate. The calculated cosmogenic ^{182}W component is ~1.7

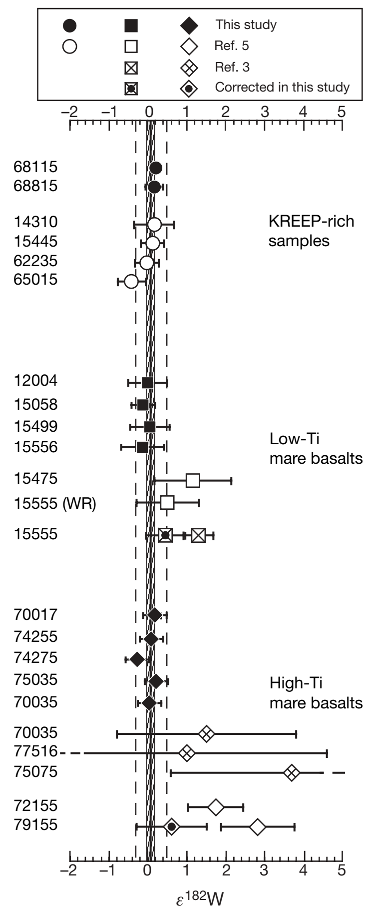


Figure 1 | ε^{182}W of lunar metals analysed in this study compared with data from refs 3 and 5. Some of the previous data (shown with black dots inside the symbols) were corrected for cosmogenic ^{182}W (see the text for details). Error bars indicate 2σ external reproducibilities. The hatched area indicates the average $\varepsilon^{182}\mathrm{W} = 0.09 \pm 0.10$ (2 s.e.m., $n = 15$) of lunar metals from this study combined with previously reported data for metals from KREEP-rich samples[5]. The dashed lines indicate a 2σ of ±0.32 ε^{182}W of these data.

[1]Institute for Isotope Geochemistry and Mineral Resources, Department of Earth Sciences, Eidgenössische Technische Hochschule Zurich, Clausiusstrasse 25, 8092 Zurich, Switzerland. [2]Institut für Mineralogie und Geochemie, Universität zu Köln, Zülpicherstrasse 49b, 50674 Köln, Germany.

Table 1 | Hf–W data for lunar metals from KREEP-rich samples, low-Ti and high-Ti mare basalts

Sample	W (p.p.m.)	Hf (p.p.m.)	Hf/W	ε^{183}W	$\varepsilon^{182}W_{meas}$	$\varepsilon^{182}W_{corr}$
KREEP-rich samples						
68115	23.9	0.407	0.02	0.00 ± 0.10	0.33 ± 0.14	0.33 ± 0.14
				−0.01 ± 0.10	0.18 ± 0.12	0.18 ± 0.12
				0.01 ± 0.09	0.18 ± 0.12	0.18 ± 0.12
				0.00 ± 0.11	0.14 ± 0.15	0.14 ± 0.15
Mean ($\pm 2\sigma$)				0.00 ± 0.02	0.21 ± 0.17	0.21 ± 0.17
68815	27.5	0.725	0.03	0.04 ± 0.11	0.02 ± 0.18	0.02 ± 0.18
				−0.02 ± 0.10	0.26 ± 0.18	0.26 ± 0.18
				0.03 ± 0.13	0.25 ± 0.15	0.25 ± 0.15
				−0.01 ± 0.13	0.22 ± 0.17	0.22 ± 0.17
Mean ($\pm 2\sigma$)				0.01 ± 0.06	0.18 ± 0.23	0.19 ± 0.23
Weighted average* ($\pm 2\sigma$; $n = 6$)				0.00 ± 0.03	0.11 ± 0.21	0.11 ± 0.21
Low-Ti mare basalts						
12004	48.7	0.539	0.11	0.04 ± 0.21	0.00 ± 0.36	0.05 ± 0.36
15058	11.3	0.113	0.01	−0.10 ± 0.15	−0.11 ± 0.19	0.01 ± 0.20
15499	7.64	0.911	0.12	0.03 ± 0.22	0.06 ± 0.31	0.16 ± 0.31
15556	21.6	0.477	0.02	0.02 ± 0.21	−0.14 ± 0.29	0.30 ± 0.36
Weighted average ($\pm 2\sigma$; $n = 4$)				−0.02 ± 0.09	−0.07 ± 0.13	0.09 ± 0.14
High-Ti mare basalts						
70017	11.6	2.18	0.19	−0.13 ± 0.14	0.20 ± 0.14	0.38 ± 0.16
70035	10.1	5.09	0.50	0.08 ± 0.13	0.05 ± 0.18	0.14 ± 0.18
74255	1.32	1.30	1.0	−0.11 ± 0.14	0.09 ± 0.17	0.11 ± 0.16
74275	8.92	2.24	0.25	−0.16 ± 0.18	−0.25 ± 0.25	−0.24 ± 0.25
75035	4.65	25.1	5.4	0.07 ± 0.15	0.22 ± 0.18	0.18 ± 0.18
Weighted average ($\pm 2\sigma$; $n = 5$)				−0.04 ± 0.14	0.11 ± 0.19	0.16 ± 0.24
Bulk lunar mantle* (2 s.e.m., $n = 15$)				0.00 ± 0.02	0.02 ± 0.09	0.09 ± 0.10

$\varepsilon^{18i}W = 10^4 \times [({}^{18i}W/{}^{184}W)_{sample}/({}^{18i}W/{}^{184}W)_{standard} - 1]$. Replicates for KREEP-rich samples are repeated measurements of the same solution. Mean values are weighted averages calculated with Isoplot (n = number of samples). Errors are $\pm 2\sigma$ unless indicated otherwise.
* Averages are calculated including data for KREEP-rich samples from ref. 3.

ε units; the elevated $^{182}W/^{184}W$ reported for 79155 is therefore presumably due mainly to cosmogenic ^{182}W. Enhanced $^{182}W/^{184}W$ in high-Ti mare basalts 72155 (ref. 3) and 75075 (ref. 5) are also not well resolved from $\varepsilon^{182}W = 0.09 \pm 0.10$ (2 s.e.m.). Details on the quantification of cosmogenic ^{182}W are given in the supplement. We conclude that in view of the identical $^{182}W/^{184}W$ we obtained for pure metal separates from nine mare basalts (average $\varepsilon^{182}W = 0.12 \pm 0.12$ (2 s.e.m.), $n = 9$), elevated $^{182}W/^{184}W$ ratios in mare basalts reported earlier probably reflect production of cosmogenic ^{182}W. This production is difficult to quantify, demonstrating that the purely radiogenic $^{182}W/^{184}W$ is difficult to determine precisely if cosmogenic ^{182}W is present. We conclude that there are no ^{182}Hf-induced tungsten isotope variations among KREEP and the mare basalt sources. Given that low-Ti and high-Ti mare basalts sample most cumulates from the lunar magma ocean (LMO)[15] and that KREEP contains a significant part of the incompatible element inventory of the lunar mantle[16], the average $\varepsilon^{182}W = 0.09 \pm 0.10$ (2 s.e.m.) of these lithologies most probably represents the average ε^{182}W of the lunar mantle. Although some ferroan anorthosites have elevated ε^{182}W values up to ~3 (ref. 6), these data are relatively imprecise and their weighted average $\varepsilon^{182}W = 1.9 \pm 1.7$ (2σ) is not resolvable from $\varepsilon^{182}W = 0.09 \pm 0.10$ (2 s.e.m.). Moreover, the analysed anorthosites might also contain cosmogenic ^{182}W.

These new constraints have far-reaching implications for the lifetime of the LMO as well as for the age and formation of the Moon. The homogeneous $^{182}W/^{184}W$ ratios of all lunar samples in spite of strongly fractionated Hf/W ratios in their source areas[17,18] indicate that the last equilibration of tungsten isotopes within the LMO occurred later than at ~60 Myr (here, Myr refers to time after formation of the first solids in the Solar System) (Fig. 2). Isotopic equilibration among the products of the LMO is possible up to a critical crystal fraction of ~60%, until which convection prevents crystal settling[19]. Although our new results only provide the earliest time (later than at ~60 Myr) for ~60% LMO crystallization, this new age is no longer in conflict with other constraints regarding the lifespan of the LMO. The rapid crystallization of the LMO required by the earlier tungsten isotope data[3,5,6] implied that some ^{147}Sm–^{143}Nd ages for ferroan anorthosites[8–11] and the ^{146}Sm–^{142}Nd model age[2,4] of the lunar mantle reflect post-LMO events[3,17]. With the revised Hf–W time constraint this is no longer required and the aforementioned Sm–Nd ages could possibly date processes associated with the LMO. This suggests that the LMO could have solidified to ~60% as late as at ~215 Myr, as given by its ^{146}Sm–^{142}Nd model age[2,4].

Our data constrain not only the lifespan of the LMO but also the timing of the giant impact. This uses the virtually identical

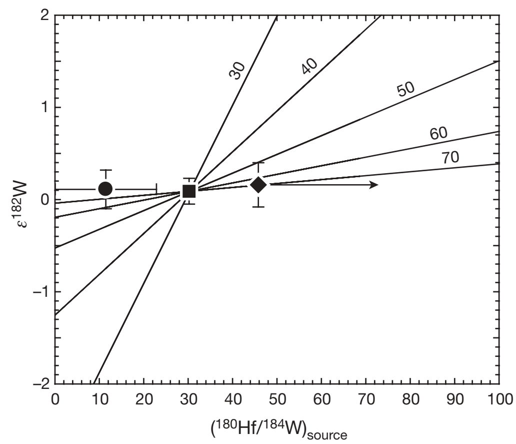


Figure 2 | Plot of ε^{182}W against source ^{180}Hf/^{184}W. Assumed Hf/W ratios are 10 ± 10 (2σ) for KREEP, 26.5 ± 1.1 (2σ) for the low-Ti mare basalt source, and >40 for the high-Ti mare basalt source. For details see Supplementary Information. ε^{182}W values shown for the low-Ti (square) and high-Ti (diamond) mare basalt sources are weighted averages of the data obtained in this study; the average ε^{182}W of KREEP (circle) was calculated by including data from ref. 3. Error bars indicate 2σ of these data. Reference isochrons corresponding to 30, 40, 50, 60 and 70 Myr, as indicated, after formation of the Solar System are shown. It is assumed that the Hf/W ratio of the bulk lunar mantle is similar to that of the low-Ti mare basalt source. The latter consists mainly of olivine and orthopyroxene, which are not capable of fractionating Hf and W (ref. 18). The identical ε^{182}W values for KREEP and the low-Ti and high-Ti mare basalt sources require that equilibration of tungsten isotopes within the LMO occurred later than ~60 Myr after the start of the Solar System. Note that this conclusion depends on neither the choice of Hf/W ratios of the sources nor on using the Hf/W ratio of the low-Ti mare basalt source as representative of the Hf/W ratio of the bulk lunar mantle.

$^{182}W/^{184}W$ ratio of the bulk silicate Moon and Earth in conjunction with their different Hf/W ratios. The latter can be inferred from U/W and Th/W ratios because U, Th and W have similar incompatibilities during mantle melting[20,21]. The lunar mantle has $U/W = 1.93 \pm 0.08$ (2σ)[21], distinctly higher than $U/W = 1.3 \pm 0.4$ (2σ) for the bulk silicate Earth (obtained from $Th/W = 5.5 \pm 1.6$ (2σ)[20] and $Th/U = 4.2 \pm 0.1$ (2σ)[22]). If refractory lithophile elements occur in chondritic proportions in planetary mantles, then $Hf/U = 13.7$ for chondrites[23] can be used to calculate $Hf/W = 26.5 \pm 1.1$ (2σ) and $Hf/W = 18.0 \pm 5.2$ (2σ) for the bulk silicate Moon and Earth. However, it is conceivable that the Moon and Earth have non-chondritic ratios of refractory lithophile elements (for example, Th/U (refs 22, 23)) and in this case the Hf/W ratios of the bulk silicate Moon and Earth might be different from those calculated above. However, this would have only a small effect on the calculated ages (Fig. 3). For instance, using a ~15% higher Hf/U ratio changes the ages by only ~2 Myr.

A first-order estimate for core formation in the Moon (or the impactor) is obtained using a two-stage model. This assumes a chondritic initial $^{182}W/^{184}W$ of the Moon and results in an age of ~37 Myr. This model implies that the bulk silicate Earth and Moon started off with slightly different $^{182}W/^{184}W$ ratios (~0.5 ε units difference at most; see Supplementary Fig. 1) and fortuitously evolved to identical present-day $^{182}W/^{184}W$ ratios. However, as the Moon consists predominantly of mantle material with high Hf/W and hence most probably radiogenic $^{182}W/^{184}W$, this two-stage model is not valid. The initial $^{182}W/^{184}W$ of the Moon was most probably higher than chondritic, resulting in an age younger than ~37 Myr and implying that the bulk silicate Moon and Earth must have had indistinguishable initial $^{182}W/^{184}W$ ratios (Supplementary Fig. 1).

A more reliable age constraint for core formation is obtained from the identical $^{182}W/^{184}W$ ratios of the bulk silicate Moon and Earth in conjunction with their distinct Hf/W ratios and by assuming identical initial $^{182}W/^{184}W$ ratios for the bulk silicate Moon and Earth. As shown in Fig. 3, core formation in the Moon must have occurred later

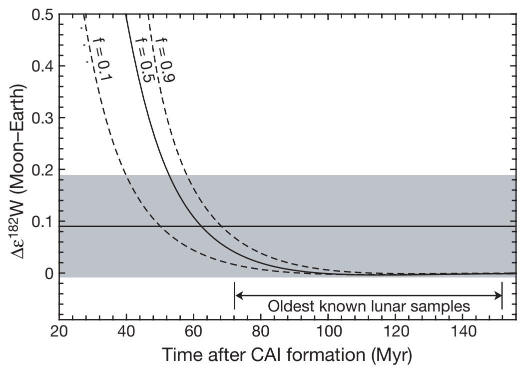


Figure 3 | ε^{182}W difference between bulk silicate Moon and Earth as a function of time of core formation. The grey shaded area indicates $\varepsilon^{182}W = 0.09 \pm 0.10$ (2 s.e.m.) for the lunar mantle as determined in this study. Assuming that the lunar and terrestrial mantles had identical initial $^{182}W/^{184}W$ ratios, the time at which these two reservoirs separated can be calculated from

$$t = \frac{1}{\lambda} \times \ln\left\{ \frac{\frac{\Delta\varepsilon^{182}W \times 10^{-4}}{(^{182}W/^{184}W)_{BSE}} \Big/ \left\{1.14 \times (Hf/U)_{CHUR} \times \left[(U/W)_{BSE} - (U/W)_{BSM}\right]\right\}}{(^{182}Hf/^{180}Hf)_0} \right\}$$

where $(^{182}Hf/^{180}Hf)_0 = 1.07 \times 10^{-4}$ (ref. 14). CHUR, chondritic uniform reservoir; BSE, bulk silicate Earth; BSM, bulk silicate Moon.

$f_{Hf/W} = \{(Hf/W)_{BSM}/(Hf/W)_{BSE} - 1\}$. Note that the lower uncertainty on the age is calculated by propagating the uncertainties on $\Delta\varepsilon^{182}W$ and $^{180}Hf/^{184}W$ and therefore does not coincide with the intersection of the $f = 0.1$ curve and the grey shaded bar. The double-headed arrow indicates the ages of the oldest known lunar samples (4.46 ± 0.04 Gyr). CAI, calcium aluminium-rich inclusion.

than at ~50 Myr, otherwise the lunar mantle would have a ^{182}W excess relative to the terrestrial mantle. This new constraint provides the earliest time that the giant impact could have occurred because the segregation of the lunar core probably occurred briefly after formation of the Moon. This also holds true if the Hf–W systematics of the Moon would largely reflect core formation in the impactor because this must pre-date the giant impact.

Because of the identical $^{182}W/^{184}W$ ratios of the bulk silicate Moon and Earth, our data provide only the earliest time at which the Moon could have formed, but the latest time is given by the age of the oldest lunar rocks. Combined with the 4.456 ± 0.040 Gyr (2σ) $^{147}Sm–^{143}Nd$ age of the lunar crust[10] we obtain an age of 62^{+90}_{-10} Myr for the formation of the Moon (Fig. 3). Given that Earth's accretion cannot have terminated before the giant impact, this also provides an age for the accretion of Earth. This is inconsistent with termination of Earth's accretion at ~30 Myr (for example, ref. 24) and is also difficult to reconcile with the <30 Myr $^{146}Sm–^{142}Nd$ model age for the differentiation of Earth's mantle[25]. Our new age constraint, however, is consistent with most U–Pb model ages for the Earth[26].

The identical $^{182}W/^{184}W$ ratios of the lunar and terrestrial mantles provide a key constraint on the formation of the Moon. The $^{182}W/^{184}W$ of any planetary mantle reflects the timescale of accretion and core formation, the degree of tungsten depletion, and the extent of re-equilibration of tungsten isotopes during core formation[24,27,28]. It is therefore unlikely that the mantles of the proto-Earth and the impactor evolved to identical $^{182}W/^{184}W$ ratios. Successful simulations of the giant impact predict that ~80% of the Moon is derived from impactor material[1], such that small pre-existing tungsten isotopic differences between proto-Earth and impactor should be reflected in the composition of the Moon. However, this is not observed in the Hf–W systematics. Similarly, the identical oxygen isotopic compositions of the Earth and Moon in spite of widespread oxygen isotopic heterogeneity among objects of the inner Solar System are unexpected. These have been interpreted to reflect accretion of the Earth and Moon from a similar mix of components formed at the same heliocentric distance[29]. As explained above, such a model cannot account for the identical tungsten isotope compositions of the lunar and terrestrial mantles. Hence, either the Moon is derived almost entirely from Earth's mantle (which is contrary to results from numerical simulations of the giant impact) or lunar and terrestrial materials equilibrated in the aftermath of the giant impact. Diffusive exchange between the silicate vapour atmosphere of the proto-Earth and the lunar magma disk might be possible for elements that became vaporized during formation of the lunar magma disk. It has been shown that this is possible for oxygen isotopes[12], but the efficiency to which tungsten became vaporized and hence could have been equilibrated isotopically remains to be investigated.

METHODS SUMMARY

Samples were crushed in an agate mortar and magnetic fractions were obtained with a hand magnet. The magnetic separates were purified by repeated grinding, magnetic separation and ultrasonication in distilled ethanol. After dissolution a ~5% aliquot was spiked with a mixed $^{180}Hf–^{183}W$ tracer for the determination of hafnium and tungsten concentrations that are used to monitor the purity of the metal separates. From the remaining ~95%, tungsten was extracted with the use of anion exchange techniques, slightly modified from refs 3, 30. All measurements were performed with a Nu Plasma MC-ICPMS at ETH Zürich. Tungsten isotope measurements were normalized to $^{186}W/^{183}W = 1.9859$ and $^{186}W/^{184}W = 0.92767$ by using the exponential law. Results obtained with these two normalization procedures agree for all samples. Tungsten isotope compositions of the samples were determined relative to the $^{182}W/^{184}W$ ratios obtained for two bracketing measurements of the ALFA AESEAR tungsten standard solution. Isobaric Os interferences on masses 186 and 184 were monitored by measuring ^{188}Os but corrections were insignificant for all samples (<0.01 p.p.m.). The accuracy of the measurements was monitored by using the $^{183}W/^{184}W$ ratio, and all samples have $^{183}W/^{184}W$ ratios identical to the standard value to within

$\pm 0.2\ \varepsilon\ (2\sigma)$ units. The external reproducibility of the tungsten isotope measurements was evaluated by repeated measurements of metals from the two KREEP-rich samples and was $\pm 0.17\ (2\sigma, n = 4)$ and $\pm 0.23\ (2\sigma, n = 4)$ (Table 1). For the low-Ti mare basalts, the reproducibility was $\sim 0.5\ \varepsilon$ units or better, and for the high-Ti mare basalts it is 0.3–0.4 ε units.

Received 12 July; accepted 24 October 2007.

1. Canup, R. M. & Asphaug, E. Origin of the Moon in a giant impact near the end of the Earth's formation. *Nature* **412,** 708–712 (2001).
2. Nyquist, L. E. *et al.* Sm-146–Nd-142 formation interval for the lunar mantle. *Geochim. Cosmochim. Acta* **59,** 2817–2837 (1995).
3. Kleine, T., Palme, H., Mezger, K. & Halliday, A. N. Hf–W chronometry of lunar metals and the age and early differentiation of the Moon. *Science* **310,** 1671–1674 (2005).
4. Rankenburg, K., Brandon, A. D. & Neal, C. R. Neodymium isotope evidence for a chondritic composition of the Moon. *Science* **312,** 1369–1372 (2006).
5. Lee, D. C., Halliday, A. N., Leya, I., Wieler, R. & Wiechert, U. Cosmogenic tungsten and the origin and earliest differentiation of the Moon. *Earth Planet. Sci. Lett.* **198,** 267–274 (2002).
6. Lee, D. C., Halliday, A. N., Snyder, G. A. & Taylor, L. A. Age and origin of the moon. *Science* **278,** 1098–1103 (1997).
7. Leya, I., Wieler, R. & Halliday, A. N. Cosmic-ray production of tungsten isotopes in lunar samples and meteorites and its implications for Hf–W cosmochemistry. *Earth Planet. Sci. Lett.* **175,** 1–12 (2000).
8. Borg, L. E. *et al.* Isotopic studies of ferroan anorthosite 62236: A young lunar crustal rock from a light rare-earth-element-depleted source. *Geochim. Cosmochim. Acta* **63,** 2679–2691 (1999).
9. Carlson, R. W. & Lugmair, G. W. The age of ferroan anorthosite 60025: oldest crust on a young Moon? *Earth Planet. Sci. Lett.* **90,** 119–130 (1988).
10. Norman, M. D., Borg, L. E., Nyquist, L. E. & Bogard, D. D. Chronology, geochemistry, and petrology of a ferroan noritic anorthosite clast from Descartes breccia 67215: Clues to the age, origin, structure, and impact history of the lunar crust. *Meteorit. Planet. Sci.* **38,** 645–661 (2003).
11. Nyquist, L. *et al.* Feldspathic clasts in Yamato-86032: Remnants of the lunar crust with implications for its formation and impact history. *Geochim. Cosmochim. Acta* **70,** 5990–6015 (2006).
12. Pahlevan, K. & Stevenson, D. J. Equilibration in the aftermath of the lunar-forming giant impact. *Earth Planet. Sci. Lett.* **262,** 438–449 (2007).
13. Leya, I., Wieler, R. & Halliday, A. N. The influence of cosmic-ray production on extinct nuclide systems. *Geochim. Cosmochim. Acta* **67,** 529–541 (2003).
14. Kleine, T., Mezger, K., Palme, H., Scherer, E. & Münker, C. Early core formation in asteroids and late accretion of chondrite parent bodies: Evidence from ^{182}Hf–^{182}W in CAIs, metal-rich chondrites and iron meteorites. *Geochim. Cosmochim. Acta* **69,** 5805–5818 (2005).
15. Shearer, C. K. & Papike, J. J. Magmatic evolution of the Moon. *Am. Mineral.* **84,** 1469–1494 (1999).
16. Palme, H. & Wänke, H. A unified trace-element model for the evolution of the lunar crust and mantle. *Proc. Lunar Sci. Conf. 6th* 1179–1202 (1975).
17. Shearer, C. K. & Newsom, H. E. W–Hf isotope abundances and the early origin and evolution of the Earth–Moon system. *Geochim. Cosmochim. Acta* **64,** 3599–3613 (2000).
18. Righter, K. & Shearer, C. K. Magmatic fractionation of Hf and W: Constraints on the timing of core formation and differentiation in the Moon and Mars. *Geochim. Cosmochim. Acta* **67,** 2497–2507 (2003).
19. Solomatov, V. S. in *Origin of the Earth and Moon* (eds Canup, R. M. & Righter, K.) 323–338 (Lunar and Planetary Institute, Houston, TX, 2000).
20. Newsom, H. E. *et al.* The depletion of W in the bulk silicate Earth: constraints on core formation. *Geochim. Cosmochim. Acta* **60,** 1155–1169 (1996).
21. Palme, H. & Rammensee, W. The significance of W in planetary differentiation processes: Evidence from new data on eucrites. *Proc. 12th Lunar Planet. Sci. Conf.* 949–964 (1981).
22. Allègre, C. J., Dupre, B. & Lewin, E. Thorium uranium ratio of the Earth. *Chem. Geol.* **56,** 219–227 (1986).
23. Rocholl, A. & Jochum, K. P. Th, U and other trace elements in carbonaceous chondrites: Implications for the terrestrial and solar-system Th/U ratios. *Earth Planet. Sci. Lett.* **117,** 265–278 (1993).
24. Jacobsen, S. B. The Hf–W isotopic system and the origin of the Earth and Moon. *Annu. Rev. Earth Planet. Sci.* **33,** 531–570 (2005).
25. Boyet, M. & Carlson, R. W. ^{142}Nd evidence for early (>4.53 Ga) global differentiation of the silicate Earth. *Science* **309,** 576–581 (2005).
26. Allègre, C. J., Manhes, G. & Gopel, C. The age of the Earth. *Geochim. Cosmochim. Acta* **59,** 1445–1456 (1995).
27. Kleine, T., Mezger, K., Palme, H., Scherer, E. & Münker, C. The W isotope evolution of the bulk silicate Earth: constraints on the timing and mechanisms of core formation and accretion. *Earth Planet. Sci. Lett.* **228,** 109–123 (2004).
28. Nimmo, F. & Agnor, C. B. Isotopic outcomes of *N*-body accretion simulations: Constraints on equilibration processes during large impacts from Hf/W observations. *Earth Planet. Sci. Lett.* **243,** 26–43 (2006).
29. Wiechert, U. *et al.* Oxygen isotopes and the moon-forming giant impact. *Science* **294,** 345–348 (2001).
30. Kleine, T., Mezger, K., Palme, H., Scherer, E. & Münker, C. The W isotope composition of eucrites metal: Constraints on the timing and cause of the thermal metamorphism of basaltic eucrites. *Earth Planet. Sci. Lett.* **231,** 41–52 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Curation and Analysis Planning Team for Extraterrestrial Materials (CAPTEM), NASA curatorial staff; G. Lofgren for supplying the Apollo lunar samples; L. Borg and A. Brandon for reviews; and F. Nimmo and J. Van Orman for discussions. This research was supported by a EU Marie Curie postdoctoral fellowship to T. Kleine.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.T. (touboul@erdw.ethz.ch).

LETTERS

Internal motions of a quasiparticle governing its ultrafast nonlinear response

P. Gaal[1], W. Kuehn[1], K. Reimann[1], M. Woerner[1], T. Elsaesser[1] & R. Hey[2]

A charged particle modifies the structure of the surrounding medium: examples include a proton in ice[1], an ion in a DNA molecule[2], an electron at an interface[3], or an electron in an organic[4] or inorganic crystal[5–7]. In turn, the medium acts back on the particle. In a polar or ionic solid, a free electron distorts the crystal lattice, displacing the atoms from their equilibrium positions. The electron, when considered together with its surrounding lattice distortion, is a single quasiparticle[5,6], known as the Fröhlich polaron[8,9]. The basic properties of polarons and their drift motion in a weak electric field are well known[10–12]. However, their nonlinear high-field properties—relevant for transport on nanometre length and ultrashort timescales—are not understood. Here we show that a high electric field in the terahertz range drives the polaron in a GaAs crystal into a highly nonlinear regime where, in addition to the drift motion, the electron is impulsively moved away from the centre of the surrounding lattice distortion. In this way, coherent lattice vibrations (phonons) and concomitant drift velocity oscillations are induced that persist for several hundred femtoseconds. They modulate the optical response at infrared frequencies between absorption and stimulated emission. Such quantum coherent processes directly affect high-frequency transport in nanostructures and may be exploited in novel terahertz-driven optical modulators and switches.

In thermal equilibrium, a Fröhlich polaron is characterized by a self-consistent attractive potential for the electron, caused by a surrounding cloud of longitudinal optical (LO) phonons. Figure 1 shows the polarized lattice (Fig. 1a) and the potential energy as a function of the relative distance between the electron and the centre of the LO phonon cloud (Fig. 1b and c). The electron–phonon coupling strength $\alpha = 0.067$ determines[13] the polaron binding energy of 5 meV, its radius of 2.7 nm (at room temperature) and its effective mass, which is slightly larger than the effective mass of the quasi-free electron.

We developed a novel experimental scheme to study the nonlinear properties of polarons in n-doped GaAs. Polarons in the crystal are accelerated by a strong electric field transient[14] with an amplitude of the order of 20 kV cm^{-1} and a duration of approximately 300 fs. The nonlinear response of the coupled electron–LO-phonon system is monitored by measuring the transmission of a mid-infrared probe pulse[15] with a centre frequency of 17 THz. The corresponding photon energy of the probe pulse is approximately twice the LO phonon energy of $\hbar\omega_{LO} = 36$ meV ($\hbar$ is Planck's constant divided by 2π, ω_{LO} is the angular frequency of the LO phonon). The probe pulse interacts with intraband transitions of polarons after a delay τ with respect to the terahertz excitation pulse. The intraband optical polarization is governed by the motion of the electronic charge in the quasiparticle and thus represents a direct probe of both the centre-of-mass motion and the internal coordinate r, describing the separation of the electron from the centre of the LO phonon cloud. In the experiment, we use the time-dependent electric field $E_{NL}(t, \tau)$ radiated from the nonlinear intraband polarization as a probe of polaron dynamics.

Both the terahertz and the probe electric field transmitted through the sample are measured in amplitude and phase as a function of real time, using electro-optic sampling[15–17]. The results presented in the following were measured at room temperature (295 K) with a 500-nm-thick Si-doped GaAs layer (doping density 10^{17} cm^{-3}) clad between 300-nm-thick $Al_{0.4}Ga_{0.6}As$ barriers[18]. Details of the experiment including the pulse generation and the detection scheme are given in the Supplementary Information.

In Fig. 2, the electric fields of the transmitted terahertz driving pulse (E_{THz}), the mid-infrared probe pulse (E_{MIR}), and the combination of the two are presented as a function of real time t for a mutual delay of the two pulses of $\tau = 77$ fs. From such transients, the change of transmission in the mid-infrared was derived:

$$\frac{\Delta T}{T_0}(t, \tau) = \frac{\int_{-\infty}^{t} E_{NL}(t', \tau) E_{MIR}(t', \tau) \mathrm{d}t'}{\int_{-\infty}^{\infty} |E_{MIR}(t', \tau)|^2 \mathrm{d}t'} \tag{1}$$

which is shown for $\tau = 77$ fs in the bottom panel of Fig. 2. In equation (1), $E_{NL}(t, \tau) = E_{both}(t, \tau) - E_{THz}(t) - E_{MIR}(t, \tau)$ is the electric field radiated from the nonlinear intraband polarization of the excited sample. The mid-infrared transmission increases in a steplike fashion while both pulses interact with the sample, and reaches a constant value after $t = 250$ fs. This demonstrates that the nonlinear increase of the mid-infrared transmission occurs only in a narrow time window.

In Fig. 3a, the electric fields $E_{THz}(t)$, $E_{MIR}(t, \tau)$, and $E_{NL}(t, \tau)$ are plotted on an extended timescale. For a delay time $\tau = 77$ fs, $E_{MIR}(t, \tau)$ and $E_{NL}(t, \tau)$ are in phase, resulting in the transmission increase shown in Fig. 2. In Fig. 3b, we plot the time-integrated transmission change $\Delta T/T_0(\infty, \tau)$ as a function of the delay time τ. This nonlinear signal shows an oscillatory behaviour with a period of 120 fs, the period of the LO phonon in GaAs. Such oscillations correspond to a periodic switching between optical gain ($\Delta T/T_0 > 0$) and absorption ($\Delta T/T_0 < 0$) on the intraband transitions probed.

We observe the oscillatory behaviour of intraband absorption and gain for THz driving fields of E_{THz} between 10 kV cm^{-1} and 30 kV cm^{-1}, the maximum field applied in our measurements. In this range, the oscillation period does not depend on E_{THz} and is always identical to the LO phonon period. For $E_{THz} < 10$ kV cm^{-1}, oscillations are absent. Experiments using a sample with a doping density five times lower do not show any nonlinear THz response. Thus, the oscillatory pump–probe signal stems exclusively from the THz excitation of electrons present by n-type doping.

The oscillatory behaviour of transmission observed here for the first time is a manifestation of the highly nonlinear response of

[1]Max-Born-Institut für Nichtlineare Optik und Kurzzeitspektroskopie, 12489 Berlin, Germany. [2]Paul-Drude-Institut für Festkörperelektronik, 10117 Berlin, Germany.

polarons to a strong external field. We analyse our findings in a nonlinear transport picture discussed in the following and described from a theoretical point of view in the Supplementary Information. An external electric field E acting on the polaron (Fig. 1d and e) induces charge transport, which is described by the drift velocity v_d. In the regime of linear response, the separation of the electron from the centre of the polaron potential is negligible (Fig. 1e) and the response to an external field is fully determined by the centre-of-mass motion of the entire quasiparticle. A weak direct current (d.c.) field (Fig. 1d) induces a dissipative drift motion along the real-space coordinate $x(t) = v_d t = \mu E t$ with a mobility $\mu = e\tau/m = 8{,}000\ \mathrm{cm^2\,V^{-1}\,s^{-1}}$ determined by the elementary charge e, the momentum relaxation time τ of the electron–phonon interaction, and the effective polaron mass m.

As the electron approaches the characteristic velocity of $v_0 = \sqrt{2\hbar\omega_{LO}/m} = 435\ \mathrm{km\,s^{-1}}$, that is, the threshold kinetic energy sufficient for emission of a LO phonon, the drift velocity v_d depends in a nonlinear way on the applied d.c. field E, as shown in Fig. 4b and discussed in refs 19–22. In this regime, the electron motion is described by the momentary electron velocity $v_e(t)$ and the differential mobility $\mu_{diff}(v_e) = [\partial E(v_e)/\partial v_e]^{-1}$. In our experimental scheme, we monitor such motion via the electric field $E_{NL}(t,\tau)$ (see equation (1)), which is radiated from the moving charge, interacting with both the driving THz field and the field $E_{MIR}(t,\tau)$ of the probe pulse. The resulting change of the mid-infrared transmission $\Delta T/T_0(\infty,\tau) \propto -\{\mu_{diff}[v_e(t)]\}^{-1}$ is determined in sign and amplitude by the inverse differential mobility.

In our experiments, the polaron potential is strongly distorted by the femtosecond terahertz field (see Fig. 2a). First, the strong external field accelerates the electron, leading to a finite distance of the electron from the centre of the polaron (along the coordinate r shown in Fig. 1g). This distance is generated impulsively—on a timescale that is short compared to the LO phonon oscillation period. As soon as the kinetic energy of the electron reaches $\hbar\omega_{LO}$, the electron velocity saturates by transferring energy to the lattice. Owing to the impulsive character of this transfer, coherent LO phonon oscillations appear as a stern wave of the moving electron (Fig. 1f), similar to wakefields in plasmas[23,24]. With increasing strength of such oscillations, the related electric field (polarization) alters the motion of the electron so that electron oscillations occur along coordinate r with a frequency ω_{LO} (Fig. 1g). On top of the drift motion of the entire quasiparticle with $v_{polaron}(t) \approx v_0$, the electron oscillations along the internal coordinate $r(t)$ are connected to a periodic modulation of the momentary electron velocity $v_e(t) = \mathrm{d}r(t)/\mathrm{d}t + v_{polaron}(t)$. In this way, the electron explores velocity regions that are characterized by different $[\mu_{diff}(v_e)]^{-1}$ of positive and negative sign (Fig. 4b), thereby modulating the transmission of the mid-infrared probe pulses in an oscillatory manner (circles in Fig. 3b). The oscillatory internal motion of the polaron is exposed to the fluctuating interaction with thermally excited LO phonons and influenced by other scattering mechanisms. Such processes result in a dephasing of the oscillations on a timescale

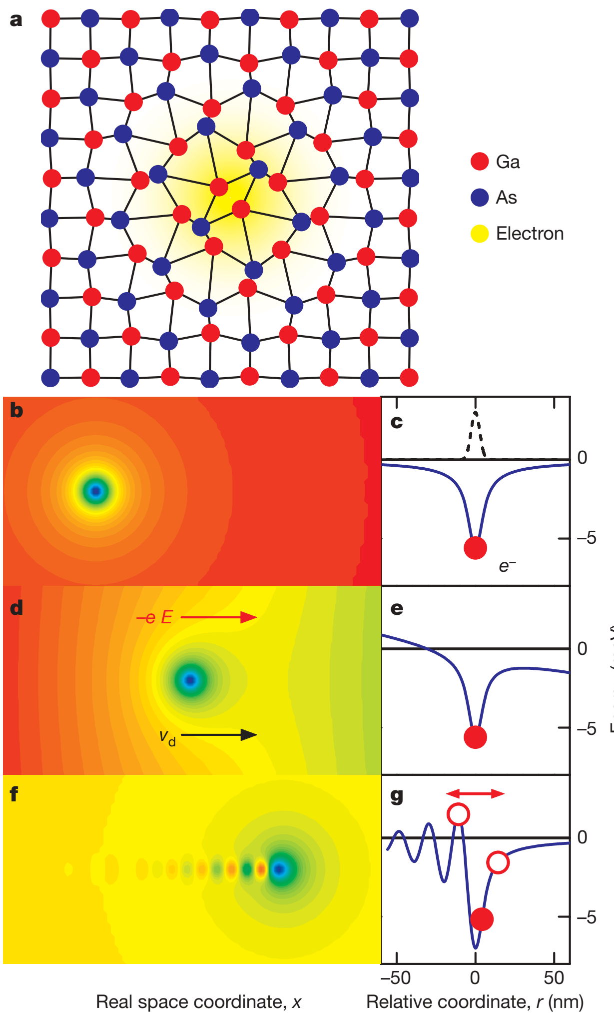


Figure 1 | The polaron. a, Lattice distortion of the Fröhlich polaron in GaAs in real space (not to scale: the actual distortion is much less and its extension is much larger). **b, c,** Self-induced polaron potential (**b,** contour plot and blue solid line in **c**) and electron wavefunction (black dashed line) of a polaron at rest self-consistently calculated with the model of ref. 13. **d, e,** Linear transport with the drift velocity v_d. For low applied electric fields E the total potential is the sum of the applied potential and the zero-field polaron potential. **f, g,** Nonlinear transport. In a strong d.c. field (which has been subtracted from the potentials shown in **f** and **g**) the drifting electron (red dot) is quasi-stationarily displaced from the minimum of the LO phonon cloud and generates coherent phonon oscillations in its stern wave. As the amplitude of coherent LO phonons exceeds a certain threshold the polaron potential eventually causes electron oscillations (shown as red circles) along the relative coordinate r on top of the drift motion of the entire quasiparticle.

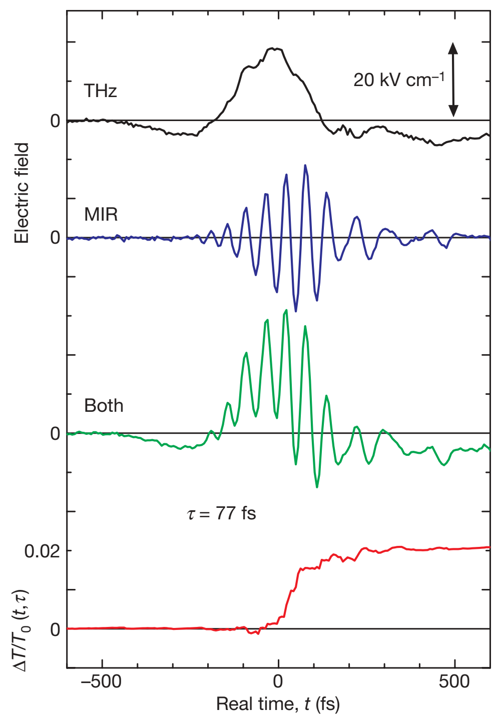


Figure 2 | Measured transients for a delay τ = 77 fs between the terahertz pump and the mid-infrared probe field. Dual-frequency chopping of both incoming beams allows for independent measurements of E_{THz}, E_{MIR}, and E_{both} (both pulses are transmitted through the sample). The red curve shows the build-up of the transmission change according to equation (1).

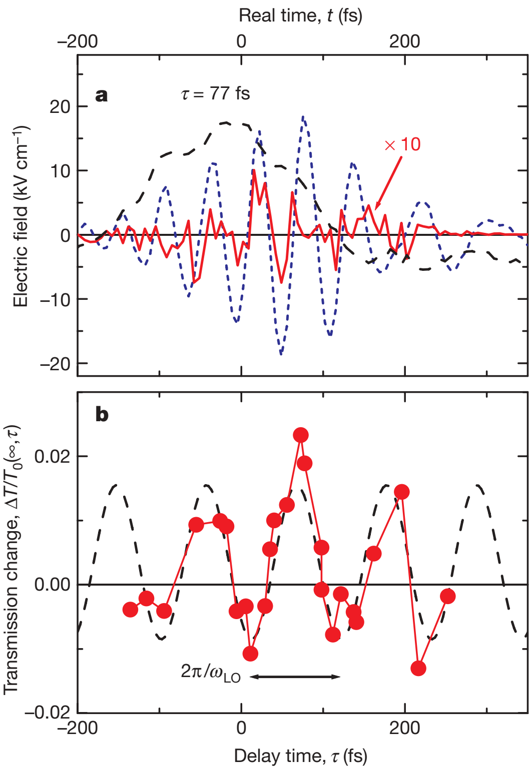


Figure 3 | Experimental results. **a**, $E_{\mathrm{THz}}(t)$ (black dashed line), $E_{\mathrm{MIR}}(t, \tau)$ (blue dotted line), and $E_{\mathrm{NL}}(t, \tau)$ (red solid line) for $\tau = 77$ fs on an extended timescale. **b**, Transmission change $\Delta T/T_0(\infty, \tau)$ of the mid-infrared pulse as a function of τ (red dots). The black dashed line shows the sine wave with the LO phonon frequency for comparison.

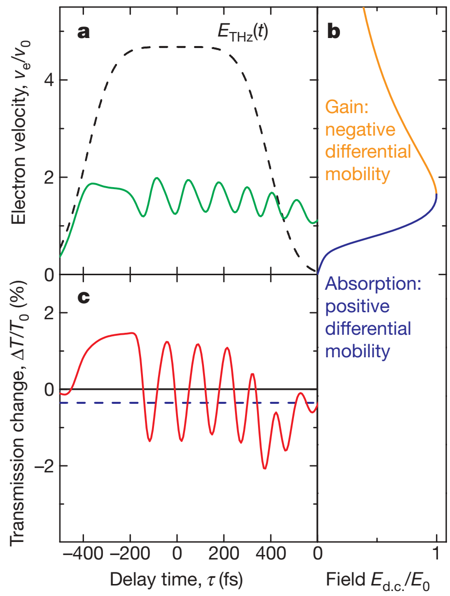


Figure 4 | Results of model calculations. **a**, Transient electron velocity (green solid line) after nonlinear excitation with a strong terahertz field (black dashed line). **b**, d.c. drift velocity v_e as a function of the applied field $E_{\mathrm{d.c.}}$. The transient modulation of the electron velocity in **a** and the mid-infrared transmission in **c** is caused by the periodic oscillation of v_e around the inflection point of the v_e–$E_{\mathrm{d.c.}}$ characteristics, changing periodically from a positive differential mobility (absorption) to a negative differential mobility (gain) in the mid-infrared spectral range. **c**, Calculated nonlinear transmission change $\Delta T/T_0(\infty, \tau)$ (red solid line) of a short mid-infrared pulse as a function of τ. The transmission reduction due to linear intraband absorption is shown for comparison (blue dashed line).

of 0.5 to 1 ps, substantially longer than the time window studied in our experiments.

The nonlinear response is analysed quantitatively by considering the time-dependent self-consistent interaction potential between the electron and the LO phonon cloud[13,25,26]. Details of the theoretical treatment are given in the Supplementary Information. The total local field acting on the electron is the sum of the externally applied fields—$E_{\mathrm{THz}}(t)$ and $E_{\mathrm{MIR}}(t, \tau)$—the field caused by the LO phonon cloud and the radiation reaction field leading to radiative damping. In Fig. 4, the time-dependent electron velocity $v_e(t)$ (green line in Fig. 4a) and the MIR transmission change (red line in Fig. 4c) calculated for a 500-fs-long terahertz pulse of $20\,\mathrm{kV\,cm^{-1}}$ amplitude (dashed line in Fig. 4a) are plotted as a function of delay time τ. Similar to the data in Fig. 3b, the nonlinear transmission change shows an oscillatory behaviour changing periodically with the frequency of the LO phonon between gain and absorption. Even the strength of the nonlinear effect, which is approximately ten times larger than the linear intraband absorption (blue dashed line in Fig. 4c), is very well reproduced by the theory. A comparison of Fig. 4a and c shows that—as expected theoretically—the oscillatory modulation of the mid-infrared transmission is directly connected to the transient motion of the electron along the internal coordinate r.

Thus, our results highlight the quantum-kinetic character of the nonlinear polaron response: the crystal lattice responds with coherent vibrations to the impulsive motion of electric charge. The timescale of such non-instantaneous process is inherently set by the LO phonon oscillation period and the picosecond decoherence of the LO phonon excitation. We note that this nonlinear phenomenon occurs at comparably low electric field amplitudes of the order of $|E| = 10\,\mathrm{kV\,cm^{-1}} = 0.1\,\mathrm{V}/(100\,\mathrm{nm})$. Thus, the quantum kinetic response is essential for high-frequency transport on nanometre length scales, in particular for highly polar materials such as GaN and Group II–Group VI semiconductors. Furthermore, the impulsive generation of coherent excitations with tailored external fields, such as phase-shaped terahertz pulses, may allow for optical modulation and switching in the femtosecond time domain.

Received 13 July; accepted 18 October 2007.

1. Cowin, J. P., Tsekouras, A. A., Iedema, M. J., Wu, K. & Ellison, G. B. Immobility of protons in ice from 30 to 190 K. *Nature* **398,** 405–407 (1999).
2. Barnett, R. N., Cleveland, C. L., Joy, A., Landman, U. & Schuster, G. B. Charge migration in DNA: ion-gated transport. *Science* **294,** 567–571 (2001).
3. Li, B. *et al.* Ultrafast interfacial proton-coupled electron transfer. *Science* **311,** 1436–1440 (2006).
4. Braun, M. *et al.* Ultrafast changes of molecular crystal structure induced by dipole solvation. *Phys. Rev. Lett.* **98,** 248301 (2007).
5. Huber, R. *et al.* How many-particle interactions develop after ultrafast excitation of an electron–hole plasma. *Nature* **414,** 286–289 (2001).
6. Hase, M., Kitajima, M., Constantinescu, A. M. & Petek, H. The birth of a quasiparticle in silicon observed in time-frequency space. *Nature* **426,** 51–54 (2003).
7. Rønnow, H. M., Renner, C., Aeppli, G., Kimura, T. & Tokura, Y. Polarons and confinement of electronic motion to two dimensions in a layered manganite. *Nature* **440,** 1025–1028 (2006).
8. Lee, T. D., Low, F. E. & Pines, D. The motion of slow electrons in a polar crystal. *Phys. Rev.* **90,** 297–302 (1953).
9. Fröhlich, H. Electrons in lattice fields. *Adv. Phys.* **3,** 325–361 (1954).
10. Sigg, H., Wyder, P. & Perenboom, J. A. A. J. Analysis of polaron effects in the cyclotron resonance of *n*-GaAs and AlGaAs-GaAs heterojunctions. *Phys. Rev. B* **31,** 5253–5261 (1985).
11. Faugeras, C. *et al.* Fröhlich mass in GaAs-based structures. *Phys. Rev. Lett.* **92,** 107403 (2004).
12. Hendry, E., Wang, F., Shan, J., Heinz, T. F. & Bonn, M. Electron transport in TiO_2 probed by THz time-domain spectroscopy. *Phys. Rev. B* **69,** 081101 (2004).
13. Peeters, F. M. & Devreese, J. T. Radius, self-induced potential, and number of virtual optical phonons of a polaron. *Phys. Rev. B* **31,** 4890–4899 (1985).
14. Bartel, T., Gaal, P., Reimann, K., Woerner, M. & Elsaesser, T. Generation of single-cycle THz transients with high electric-field amplitudes. *Opt. Lett.* **30,** 2805–2807 (2005).
15. Reimann, K., Smith, R. P., Weiner, A. M., Elsaesser, T. & Woerner, M. Direct field-resolved detection of terahertz transients with amplitudes of megavolts per centimeter. *Opt. Lett.* **28,** 471–473 (2003).
16. Wu, Q. & Zhang, X.-C. Free-space electro-optic sampling of terahertz beams. *Appl. Phys. Lett.* **67,** 3523–3525 (1995).

17. Wu, Q. & Zhang, X.-C. Free-space electro-optic sampling of mid-infrared pulses. *Appl. Phys. Lett.* **71,** 1285–1286 (1997).
18. Gaal, P. *et al.* Nonlinear terahertz response of *n*-type GaAs. *Phys. Rev. Lett.* **96,** 187402 (2006).
19. Bányai, L. Motion of a classical polaron in a dc electric field. *Phys. Rev. Lett.* **70,** 1674–1677 (1993).
20. Janssen, N. & Zwerger, W. Nonlinear transport of polarons. *Phys. Rev. B* **52,** 9406–9417 (1995).
21. Jensen, J. H. & Sauls, J. A. Polarons near the Čerenkov velocity. *Phys. Rev. B* **38,** 13387–13394 (1988).
22. Meinert, G., Bányai, L. & Gartner, P. Classical polarons in a constant electric field. *Phys. Rev. B* **63,** 245203 (2001).
23. Bingham, R. On the crest of a wake. *Nature* **445,** 721–722 (2007).
24. Blumenfeld, I. *et al.* Energy doubling of 42 GeV electrons in a metre-scale plasma wakefield accelerator. *Nature* **445,** 741–744 (2007).
25. Magnus, W. & Schoenmaker, W. Dissipative motion of an electron-phonon system in a uniform electric field: An exact solution. *Phys. Rev. B* **47,** 1276–1281 (1993).
26. Stroucken, T., Knorr, A., Thomas, P. & Koch, S. W. Coherent dynamics of radiatively coupled quantum well excitons. *Phys. Rev. B* **53,** 2026–2033 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge financial support by the Deutsche Forschungsgemeinschaft.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.W. (woerner@mbi-berlin.de).

nature Vol 450|20/27 December 2007|doi:10.1038/nature06401

LETTERS

Optical frequency comb generation from a monolithic microresonator

P. Del'Haye[1], A. Schliesser[1], O. Arcizet[1], T. Wilken[1], R. Holzwarth[1] & T. J. Kippenberg[1]

Optical frequency combs[1–3] provide equidistant frequency markers in the infrared, visible and ultraviolet[4,5], and can be used to link an unknown optical frequency to a radio or microwave frequency reference[6,7]. Since their inception, frequency combs have triggered substantial advances in optical frequency metrology and precision measurements[6,7] and in applications such as broadband laser-based gas sensing[8] and molecular fingerprinting[9]. Early work generated frequency combs by intra-cavity phase modulation[10,11]; subsequently, frequency combs have been generated using the comb-like mode structure of mode-locked lasers, whose repetition rate and carrier envelope phase can be stabilized[12]. Here we report a substantially different approach to comb generation, in which equally spaced frequency markers are produced by the interaction between a continuous-wave pump laser of a known frequency with the modes of a monolithic ultra-high-*Q* microresonator[13] via the Kerr nonlinearity[14,15]. The intrinsically broadband nature of parametric gain makes it possible to generate discrete comb modes over a 500-nm-wide span (~70 THz) around 1,550 nm without relying on any external spectral broadening. Optical-heterodyne-based measurements reveal that cascaded parametric interactions give rise to an optical frequency comb, overcoming passive cavity dispersion. The uniformity of the mode spacing has been verified to within a relative experimental precision of 7.3×10^{-18}. In contrast to femtosecond mode-locked lasers[16], this work represents a step towards a monolithic optical frequency comb generator, allowing considerable reduction in size, complexity and power consumption. Moreover, the approach can operate at previously unattainable repetition rates[17], exceeding 100 GHz, which are useful in applications where access to individual comb modes is required, such as optical waveform synthesis[18], high capacity telecommunications or astrophysical spectrometer calibration[19].

Optical microcavities[20] are, owing to their long temporal and small spatial light confinement, ideally suited for nonlinear frequency conversion, which has led to a marked reduction in the threshold of nonlinear optical processes[21–23]. In contrast to stimulated gain, parametric gain[24] does not involve coupling to a dissipative reservoir, is broadband as it does not rely on atomic or molecular resonances, and constitutes a phase-sensitive amplification process, making it uniquely suited for tunable frequency conversion. In the case of a material with inversion symmetry—such as silica—the nonlinear optical effects are dominated by the third-order nonlinearity. The parametric process is based on four-wave mixing (FWM) among two pump photons (frequency υ_P) with a signal (υ_S) and idler photon (υ_I) and results in the emergence of (phase-coherent) signal and idler sideband) at the expense of the pump. Observing parametric interactions requires both energy and momentum conservation to be satisfied. In a microcavity of the whispering gallery type[20] the optical modes are angular momentum eigenstates and have discrete propagation constants $\beta_m = m/R$ resulting from the periodic boundary condition, where the integer m denotes the angular mode number and R the cavity radius. Consequently, the conversion of two pump photons (propagation constant β_N) into adjacent signal and idler modes ($\beta_{N-\Delta N}$, $\beta_{N+\Delta N}$, $\Delta N = 1, 2, 3\ldots$) conserves momentum intrinsically[14]. On the other hand, energy conservation ($2h\upsilon_P = h\upsilon_I + h\upsilon_S$, where h is the Planck constant) places stringent conditions on cavity dispersion, because the frequency difference between adjacent modes $\upsilon_{FSR} = |\upsilon_m - \upsilon_{m+1}|$ (the free spectral range, FSR) can vary owing to both material and geometric dispersion, rendering the cavity modes non-equidistant. For modes sufficiently close to the pump, however, the accumulated variation in the FSRs can be small. Indeed under these circumstances, parametric frequency conversion has only recently been observed for the first time[14,15] in ultra-high-Q microcavities (made of crystalline[15] CaF_2 and silica[14,30]).

Importantly, this mechanism can also be used to generate optical frequency combs: the initially generated signal and idler sidebands can interact among each other and produce higher-order sidebands (Fig. 1) by non-degenerate FWM[26] which ensures that the frequency difference of pump and first-order sidebands $\Delta\upsilon \equiv |\upsilon_P - \upsilon_S| = |\upsilon_P - \upsilon_I|$ is exactly transferred to all higher-order sidebands. This can be readily seen by noting that, for example, the second-order sidebands are generated by mixing among the pump and first-order signal/idler sidebands (for example, $\upsilon_{S2} = \upsilon_P + \upsilon_S - \upsilon_I = \upsilon_P - 2\Delta\upsilon$), which rigidly determines the spacing of any successively higher sidebands (Fig. 1b). Thus, provided that the cavity shows low dispersion, the successive FWM to higher orders would intrinsically lead to the generation of phase-coherent sidebands with equal spacing, that is, an optical frequency comb. Note that the generation of an optical frequency comb in this fashion requires experimental verification, as the parametric process itself could in principle produce signal/idler sidebands that are only pair-wise equidistant (for example, through degenerate FWM to higher orders) but not mutually equidistant as required for comb generation. Here, we report that monolithic microresonators do indeed allow realization of massively cascaded sideband generation, whose perfect equidistance is confirmed to a level of 7.3×10^{-18} (when normalized to the optical carrier). We have thus demonstrated optical frequency comb generation from a continuously pumped microcavity.

We use ultra-high-Q monolithic microresonators in the form of silica toroidal microcavities[13] on a silicon chip, which possess long photon storage times (τ_0): that is, ultra-high quality factors ($Q = 2\pi\upsilon\tau_0 > 10^8$) and small mode volumes. Highly efficient coupling is achieved using tapered optical fibres[27]. Owing to the high circulating power, parametric interactions are readily observed at a threshold of about 50 μW. When pumping with a continuous-wave 1,550-nm laser source, we observe a massive cascade and multiplication of the parametric sidebands extending to both higher and lower

[1]Max Planck Institut für Quantenoptik (MPQ), Hans-Kopfermann-Strasse 1, 85748 Garching, Germany.

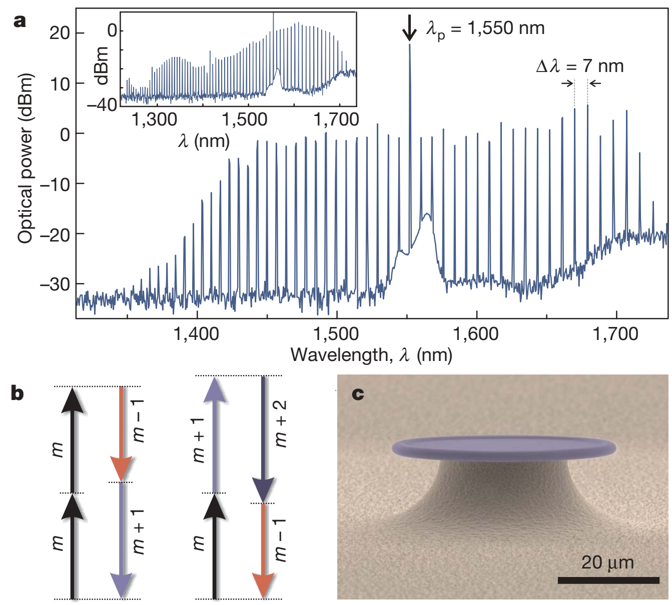


Figure 1 | Broadband parametric frequency conversion from a monolithic toroidal microresonator. **a**, Spectrum of the parametric frequency conversion observed in a 75-µm-diameter monolithic toroid microcavity when pumped with 60-mW continuous-wave laser power at 1,550 nm. The combination of parametric interactions and FWM gives rise to a broadband emission, spaced by the cavity free spectral range. Units of dBm represent logarithmic power relative to power of 1 mW. Inset: broadband parametric conversion of a different sample generating more than 70 parametric modes extending over a wavelength span of nearly 500 nm (pump power 130 mW). The asymmetry in the spectrum (with higher power in the higher-wavelength sidebands) is attributed to Raman amplification and variation of the tapered fibre output coupling. **b**, Schematic of the processes that contribute to the parametric conversion: degenerate (left) and non-degenerate (right) FWM among cavity eigenmodes. Momentum conservation is intrinsically satisfied for the designated angular mode number (m) combinations. **c**, Scanning electron microscope image of a toroid microcavity on a silicon chip.

frequencies. Figure 1a shows a spectrum in which a 75-µm-diameter microcavity was pumped with 60 mW power, giving rise to an intracavity intensity exceeding 1 GW cm^{-2}. The parametric frequency conversion could extend over more than 490 nm (see Fig. 1a inset), with the total conversion efficiency being 21.2%. These bright sidebands (termed Kerr combs in the remaining discussions) could be observed in many different samples, with conversion efficiencies of more than 83% by working in the over-coupled regime. In the largest fabricated samples (190 µm diameter), 380-nm broad Kerr combs comprising 134 modes spaced by 375 GHz could be generated at the expense of slightly higher pump power (see Supplementary Information).

To verify that the Kerr comb contains equidistant frequencies, we used a fibre-laser-based optical frequency comb[28] from Menlo Systems (termed reference comb in the remaining discussion) as a reference grid with a repetition rate of f_{rep} = 100 MHz. The principle underlying our measurement is that the beating generated on a photodiode by superimposing the reference comb with the Kerr comb will produce beat notes that constitute a replica of the optical spectrum in the radio frequency domain, similar to multi-heterodyne frequency comb spectroscopy[29] (see Supplementary Information). Specifically, if the Kerr comb is equidistant, the beat notes with the reference comb will constitute an equidistant comb in the radio frequency domain (with frequency spacing Δf, where $\Delta f = (\Delta\upsilon \bmod f_{rep})$). Figure 2a shows the experimental setup for the optical beat measurement. In brief, an external cavity laser operating at 1,550 nm was used to pump a microcavity. The generated Kerr modes were then superimposed with the reference comb in a beat note detection unit (BDU), consisting of polarizing optics for combining reference and Kerr comb and a grating to select the desired region of spectral overlap. In this manner, the beatings of nine simultaneously oscillating parametric modes (covering a wavelength span of more than 50 nm) were recorded, as shown in Fig. 2c. Remarkably, from the equidistant spacing of the radio frequencies, it is found that the generated sidebands are equidistant to within 5 kHz (as determined by the measurement time of 200 µs).

To improve the accuracy, we developed an additional experiment measuring the beat notes of three Kerr modes with the fibre-reference comb using three separate BDUs (Fig. 3a), each of which counted a single radio frequency beat (f_0, f_1, f_2). A signal-to-noise ratio exceeding 30 dB in a 500-kHz bandwidth was achieved, sufficient to use radio frequency counters, all referenced to a 10-MHz signal provided by a hydrogen maser. The beat note measured on BDU1 was used to

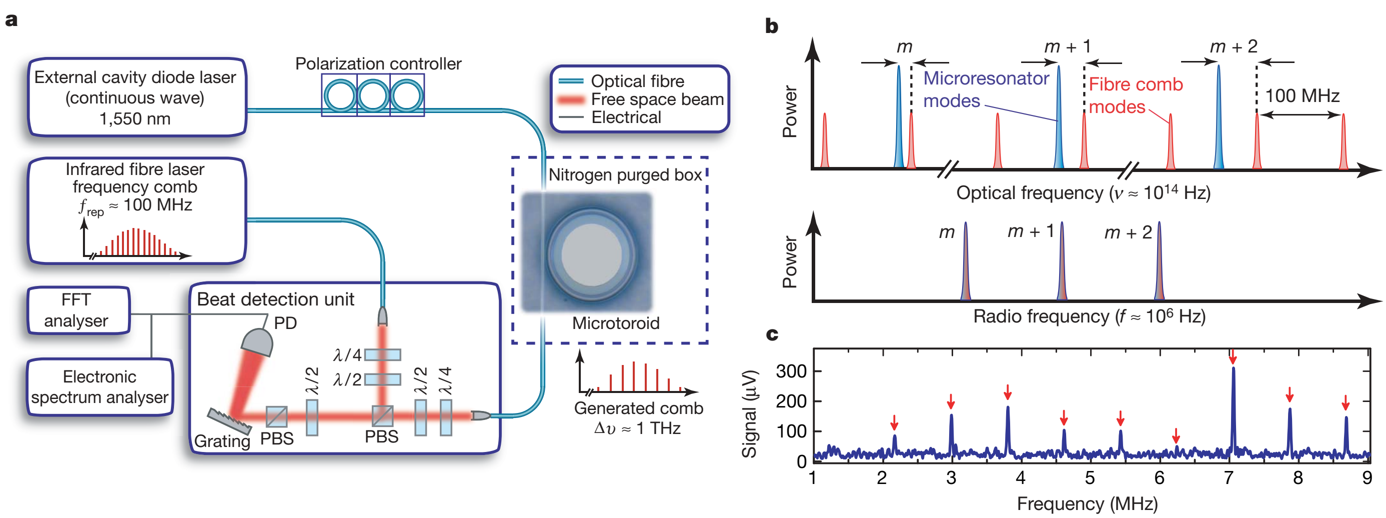


Figure 2 | Multi-heterodyne beat note detection. **a**, The experimental setup consisting of an external cavity laser (continuous wave) coupled to an ultra-high-Q monolithic microresonator in a nitrogen environment via a tapered fibre. The parametric output is coupled into one port of a BDU. The second port of the BDU is coupled to a mode-locked femtosecond erbium-doped fibre laser that serves as a reference comb. A grating is used to select a spectral region of the Kerr comb modes and a p-i-n silicon photodiode (PD) records their beatings with the reference comb (see Supplementary Information for details). PBS, polarizing beamsplitter; FFT, fast Fourier transform analyser. **b**, The measurement principle. The beating of the reference comb with the microcavity parametric modes yields beat frequencies in the radio frequency domain. **c**, Radio frequency spectrum of nine simultaneously oscillating Kerr modes, showing a uniform spacing.

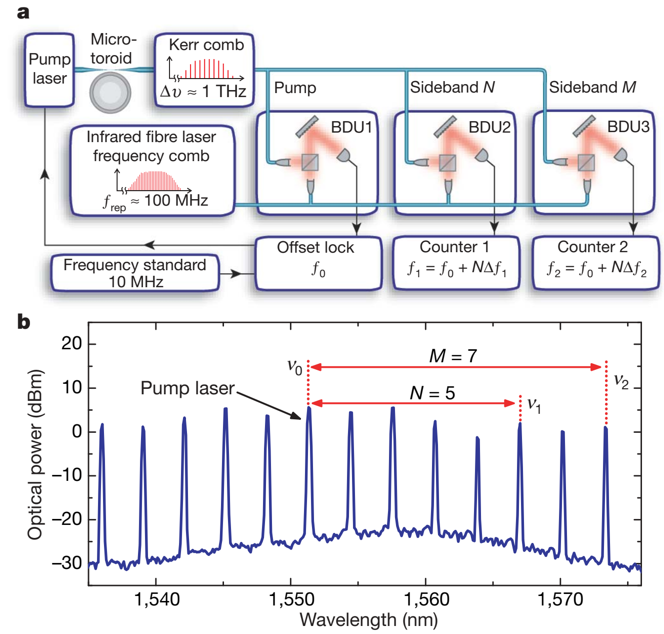


Figure 3 | Probing the equidistance of the comb structure. a, Simplified schematic of the setup which consists of three BDUs to measure the beating of three Kerr modes with the infrared fibre-based reference comb. All BDUs are referenced to the MPQ hydrogen maser as the frequency standard. The first BDU is used to produce a phase lock between one comb line of the reference comb and the pump laser (which constitutes one mode of the Kerr comb). **b**, The parametric spectrum used to validate the equidistance of the comb modes.

create an offset lock between a single reference comb mode and the pump laser by a known offset frequency (f_0). The second (third) counter measured the Nth (Mth) mode of the Kerr comb as shown in Fig. 3. For equidistant mode spacing, the second (third) BDU gives rise to the beat frequency $f_1 = f_0 + \Delta f \times N$ ($f_2 = f_0 + \Delta f \times M$). We then checked the uniformity of the Kerr comb by deriving the deviation from equidistant mode spacing, that is, $\varepsilon \equiv \frac{f_2 - f_1}{M - N} - \frac{f_1 - f_0}{N}$.

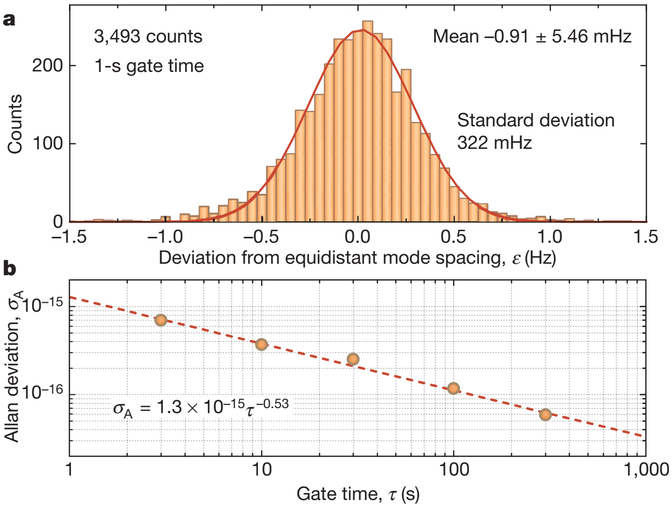


Figure 4 | Verification of the equidistant mode spacing. a, The deviation from equidistant mode spacing (ε) for a gate time of 1 s for the parametric spectrum and measurement setup described in Fig. 3. For this measurement 3,493 points were recorded. The solid red line is a gaussian fit to the distribution. The standard deviation of the mean implies an accuracy of the mode spacing at the millihertz level, confirming the comb-like structure of the parametric spectrum. **b**, Allan deviation as a function of gate time, showing an inverse square-root dependence, as determined by the fit (dashed red line).

Alternatively, the ratio $(f_1 - f_0)/(f_2 - f_0) = N/M$ was counted directly (using frequency mixing and ratio counting; see Supplementary Information). Figure 4a shows the measured gaussian distribution of ε for $N = 5$, $M = 7$ and a counter gate time τ of 1 s. The cavity modes of this measurement span 21 nm and yield a deviation from equidistance of $\varepsilon = (-0.9 \pm 5.5)$ mHz. The wavelength span covered by this measurement was limited by the gain bandwidth of an erbium doped fibre amplifier, which was necessary for amplification of the reference comb to ensure sufficient power to run three BDUs simultaneously at the required signal-to-noise ratio. The Allan deviation of ε, measured for several gate times, is reported in Fig. 4b. Table 1 shows the results for different gate times and for the two different counting methods (the complete list is contained in the Supplementary Information). The weighted average of these results verifies the uniformity of the comb spacing to a level of 7.3×10^{-18}, when referenced to the optical carrier. Normalized to the bandwidth of the measured Kerr lines (2.1 THz), this corresponds to 5.2×10^{-16}. This accuracy is on a par with measurements for fibre-based frequency combs[28] and confirms that the generated Kerr combs show uniform mode spacing.

Next we investigated the role of dispersion underlying the observed comb generation. Dispersion in whispering-gallery-mode microcavities is characterized by the variation in the FSR, $\Delta\upsilon_{\mathrm{FSR}} = (\upsilon_{m+1} - \upsilon_m) - (\upsilon_m - \upsilon_{m-1}) = \upsilon_{m+1} + \upsilon_{m-1} - 2\upsilon_m$, and has two contributions. Geometrical cavity dispersion accounts for a negative FSR dispersion, given by $\Delta\upsilon_{\mathrm{FSR}} \approx -0.41 \frac{c}{2\pi n R} m^{-\frac{5}{3}}$ where R is the cavity radius (see Supplementary Information). Material dispersion on the other hand is given by $\Delta\upsilon_{\mathrm{FSR}} \approx \frac{1}{4\pi^2} \frac{c^2 \lambda^2}{n^3 R^2} \cdot \mathrm{GVD}$, where $\mathrm{GVD} = -\frac{\lambda}{c} \frac{\partial^2 n}{\partial \lambda^2}$ is the group velocity dispersion parameter. Because the GVD of silica is positive for wavelength greater than 1.3 μm (anomalous dispersion), it can compensate[30] the intrinsic resonator dispersion (causing $\Delta\upsilon_{\mathrm{FSR}} > 0$). Indeed we measured a positive dispersion (see Supplementary Information) which equates to only about 20 MHz over a span of about 60 nm. This low value indicates that the experiments are carried out close to the zero dispersion wavelength, in agreement with theoretical predictions.

Note that the residual cavity dispersion exceeding the 'cold' cavity linewidth does not preclude the parametric comb generation process. This could be explained in terms of a nonlinear optical mode pulling process as reported in ref. 14. The strong continuous-wave pump laser will induce both self-phase modulation and cross-phase modulation, the latter being twice as large as the former. The refractive index changes induced by self-phase and cross-phase modulation will shift the cavity resonance frequencies by different amounts, thereby causing a net change in the (driven) cavity dispersion from its passive (undriven) value[14]. This nonlinear mode pulling can provide a mechanism to compensate the residual cavity dispersion.

Regarding future applications in metrology, we note that absolute referencing can be attained by locking the pump laser frequency υ_0 to a known atomic transition and locking the mode spacing to a microwave reference (such as a caesium atomic clock). The latter requires the two degrees of freedom of the comb, its repetition rate (that is,

Table 1 | Experimental results on the deviation from equidistant mode spacing

Gate time (s)	Number of readings	Mean ε (mHz)	Standard deviation (mHz)	Relative accuracy	Counting method
1	3,493	-0.9 ± 5.5	322	2.7×10^{-17}	Two counters
3	173	5.8 ± 12.6	165	6.3×10^{-17}	Ratio counting
10	22	-17.9 ± 15.0	70	7.5×10^{-17}	Ratio counting
30	39	1.7 ± 7.4	46	3.7×10^{-17}	Ratio counting
100	42	-0.3 ± 2.7	17	1.4×10^{-17}	Ratio counting
300	14	-0.8 ± 2.8	11	1.4×10^{-17}	Ratio counting

The weighted mean of all measurements (including Supplementary Information) yields a relative accuracy of 7.3×10^{-18}.

mode spacing, $\Delta\upsilon$) and frequency offset (that is, $\upsilon_{CEO} = \upsilon_0 \bmod \Delta\upsilon$), to be controlled independently. We have already shown in a proof-of-concept experiment that it is possible to lock two modes of the Kerr comb simultaneously to two modes of the reference comb, showing that independent control of both υ_{CEO} and $\Delta\upsilon$ is possible. The two actuators used for this lock are the detuning of the pump laser from the microcavity resonance and the pump power, which affects the optical pathlength of the cavity via the thermal effect and the nonlinear phase shift.

A monolithic frequency comb generator could potentially prove useful for frequency metrology, given further improvements. Evidently a readily measurable repetition rate would be highly advantageous when directly referencing the optical field to a microwave signal[2]. To this end a 660-μm-diameter microcavity would already allow operating at repetition rates less than 100 GHz, which can be detected using fast photodiodes. On the other hand, a large mode spacing as demonstrated here could prove useful in applications such as line-by-line pulse shaping[18], calibration of astrophysical spectrometers[19] or direct comb spectroscopy. The high repetition rate from an on-chip device may also be useful in generating multiple channels for high capacity telecommunications (spacing 160 GHz) and low-noise microwave signals. Furthermore, we note that parametric interactions do also occur in other types of microcavities—for example, CaF_2 (ref. 15)—provided that the material shows a third-order nonlinearity and sufficiently long photon lifetimes. Therefore the cavity geometry is not conceptually central to the work and the reported phenomena should be observable in other types of high-Q microresonators, such as silicon, silicon-on-insulator or crystalline-based whispering-gallery-mode resonators. The recent observation of net parametric gain[25] on a silicon chip is a promising step in this direction.

Received 3 April; accepted 12 October 2007.

1. Udem, T., Holzwarth, R. & Hänsch, T. W. Optical frequency metrology. *Nature* **416**, 233–237 (2002).
2. Cundiff, S. T. & Ye, J. Colloquium: Femtosecond optical frequency combs. *Rev. Mod. Phys.* **75**, 325–342 (2003).
3. Ye, J. & Cundiff, S. T. *Femtosecond Optical Frequency Comb: Principle, Operation and Applications* (Springer, New York, 2005).
4. Jones, R. J., Moll, K. D., Thorpe, M. J. & Ye, J. Phase-coherent frequency combs in the vacuum ultraviolet via high-harmonic generation inside a femtosecond enhancement cavity. *Phys. Rev. Lett.* **94**, 193201 (2005).
5. Gohle, C. *et al.* A frequency comb in the extreme ultraviolet. *Nature* **436**, 234–237 (2005).
6. Diddams, S. A. *et al.* Direct link between microwave and optical frequencies with a 300 THz femtosecond laser comb. *Phys. Rev. Lett.* **84**, 5102–5105 (2000).
7. Reichert, J. *et al.* Phase coherent vacuum-ultraviolet to radio frequency comparison with a mode-locked laser. *Phys. Rev. Lett.* **84**, 3232–3235 (2000).
8. Thorpe, M. J., Moll, K. D., Jones, J. J., Safdi, B. & Ye, J. Broadband cavity ringdown spectroscopy for sensitive and rapid molecular detection. *Science* **311**, 1595–1599 (2006).
9. Diddams, S. A., Hollberg, L. & Mbele, V. Molecular fingerprinting with the resolved modes of a femtosecond laser frequency comb. *Nature* **445**, 627–630 (2007).
10. Kourogi, M., Nakagawa, K. & Ohtsu, M. Wide-span optical frequency comb generator for accurate optical frequency difference measurement. *IEEE J. Quantum Electron.* **29**, 2693–2701 (1993).
11. Ye, J., Ma, L. S., Daly, T. & Hall, J. L. Highly selective terahertz optical frequency comb generator. *Opt. Lett.* **22**, 301–303 (1997).
12. Jones, D. J. *et al.* Carrier-envelope phase control of femtosecond mode-locked lasers and direct optical frequency synthesis. *Science* **288**, 635–639 (2000).
13. Armani, D. K., Kippenberg, T. J., Spillane, S. M. & Vahala, K. J. Ultra-high-*Q* toroid microcavity on a chip. *Nature* **421**, 925–928 (2003).
14. Kippenberg, T. J., Spillane, S. M. & Vahala, K. J. Kerr-nonlinearity optical parametric oscillation in an ultrahigh-Q toroid microcavity. *Phys. Rev. Lett.* **93**, 083904 (2004).
15. Savchenkov, A. A. *et al.* Low threshold optical oscillations in a whispering gallery mode CaF_2 resonator. *Phys. Rev. Lett.* **93**, 243905 (2004).
16. Steinmeyer, G., Sutter, D. H., Gallmann, L., Matuschek, N. & Keller, U. Frontiers in ultrashort pulse generation: Pushing the limits in linear and nonlinear optics. *Science* **286**, 1507–1512 (1999).
17. Keller, U. Recent developments in compact ultrafast lasers. *Nature* **424**, 831–838 (2003).
18. Weiner, A. M. Femtosecond pulse shaping using spatial light modulators. *Rev. Sci. Instrum.* **71**, 1929–1960 (2000).
19. Murphy, M. T. *et al.* High-precision wavelength calibration with laser frequency combs. *Mon. Not. R. Astron. Soc.* **380**, 839–847 (2007).
20. Vahala, K. J. Optical microcavities. *Nature* **424**, 839–846 (2003).
21. Chang, R. K. & Campillo, A. J. *Optical Processes in Microcavities* (World Scientific, Singapore, 1996).
22. Spillane, S. M., Kippenberg, T. J. & Vahala, K. J. Ultralow-threshold Raman laser using a spherical dielectric microcavity. *Nature* **415**, 621–623 (2002).
23. Carmon, T. & Vahala, K. J. Visible continuous emission from a silica microphotonic device by the third harmonic generation. *Nature Phys.* **3**, 430–435 (2007).
24. Dunn, M. H. & Ebrahimzadeh, M. Parametric generation of tunable light from continuous-wave to femtosecond pulses. *Science* **286**, 1513–1517 (1999).
25. Foster, M. A. *et al.* Broad-band optical parametric gain on a silicon photonic chip. *Nature* **441**, 960–963 (2006).
26. Stolen, R. H. & Bjorkholm, J. E. Parametric amplification and frequency-conversion in optical fibers. *IEEE J. Quantum Electron.* **18**, 1062–1072 (1982).
27. Spillane, S. M., Kippenberg, T. J., Painter, O. J. & Vahala, K. J. Ideality in a fiber-taper-coupled microresonator system for application to cavity quantum electrodynamics. *Phys. Rev. Lett.* **91**, 043902 (2003).
28. Kubina, P. *et al.* Long term comparison of two fiber based frequency comb systems. *Opt. Expr.* **13**, 904–909 (2005).
29. Schliesser, A., Brehm, M., Keilmann, F. & van der Weide, D. W. Frequency-comb infrared spectrometer for rapid, remote chemical sensing. *Opt. Expr.* **13**, 9029–9038 (2005).
30. Agha, I. H., Okawachi, Y., Foster, M. A., Sharping, J. E. & Gaeta, A. L. Four-wave mixing parametric oscillations in dispersion-compensated high-Q silica microspheres. *Phys. Rev. A* **76**, 043837 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. W. Hänsch, T. Udem, K. J. Vahala and S. A. Diddams for critical discussions and suggestions. T.J.K. acknowledges support via an Independent Max Planck Junior Research Group. This work was funded as part of a Marie Curie Excellence Grant (RG-UHQ), the DFG funded Nanosystems Initiative Munich (NIM) and a Marie Curie Reintegration Grant (JRG-UHQ). We thank J. Kotthaus for access to clean-room facilities for sample fabrication.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to T.J.K. (tobias.kippenberg@mpq.mpg.de).

LETTERS

Environmental precursors to rapid light carbon injection at the Palaeocene/Eocene boundary

Appy Sluijs[1], Henk Brinkhuis[1], Stefan Schouten[3], Steven M. Bohaty[4], Cédric M. John[4]†, James C. Zachos[4], Gert-Jan Reichart[2], Jaap S. Sinninghe Damsté[2,3], Erica M. Crouch[1]† & Gerald R. Dickens[5]

The start of the Palaeocene/Eocene thermal maximum—a period of exceptional global warming about 55 million years ago—is marked by a prominent negative carbon isotope excursion that reflects a massive input of ^{13}C-depleted ('light') carbon to the ocean–atmosphere system[1]. It is often assumed[2] that this carbon injection initiated the rapid increase in global surface temperatures and environmental change that characterize the climate perturbation[3–7], but the exact sequence of events remains uncertain. Here we present chemical and biotic records of environmental change across the Palaeocene/Eocene boundary from two sediment sections in New Jersey that have high sediment accumulation rates. We show that the onsets of environmental change (as recorded by the abundant occurrence ('acme') of the dinoflagellate cyst *Apectodinium*) and of surface-ocean warming (as evidenced by the palaeothermometer TEX_{86}) preceded the light carbon injection by several thousand years. The onset of the *Apectodinium* acme also precedes the carbon isotope excursion in sections from the southwest Pacific Ocean[8] and the North Sea, indicating that the early onset of environmental change was not confined to the New Jersey shelf. The lag of ~3,000 years between the onset of warming in New Jersey shelf waters and the carbon isotope excursion is consistent with the hypothesis that bottom water warming caused the injection of ^{13}C-depleted carbon by triggering the dissociation of submarine methane hydrates[1,9,10], but the cause of the early warming remains uncertain.

The globally recognized carbon isotope excursion (CIE) at the Palaeocene/Eocene thermal maximum (PETM) marks a >2.5‰ decrease in the stable carbon isotope composition (δ^{13}C) of sedimentary components[3,11]. It signifies an enormous (at least 1.5×10^{18} g) input of ^{13}C-depleted carbon to the ocean and atmosphere, analogous in both magnitude and composition to current and expected fossil fuel emissions[1,2,12]. One prominent example of biotic change associated with the onset of the CIE is recorded along continental margins, where sediment sequences from all latitudes contain high abundances of dinoflagellate cysts (dinocysts; Supplementary Information) belonging to the subtropical genus *Apectodinium*[6–8,13]; the *Apectodinium* acme. The source and impact of the ^{13}C-depleted carbon input, however, remain controversial[1,2,12,14]. One possibility is the rapid dissociation of submarine methane hydrates[1]; other theories include the injection of thermogenic methane[15] and oxidation of sedimentary organic carbon[12]. Underlying the current debate is whether the input of ^{13}C-depleted carbon to the atmosphere caused or resulted from global warming and environmental change.

Unravelling the sequence of events at the onset of the PETM has been difficult, given the rapid nature of the perturbation. Some climate proxy records suggest that onset of warming slightly preceded the CIE[5,16]. However, these records are from slowly accumulating deep marine sediment sections, where bioturbation has blurred various signals and carbonate dissolution reduced temporal resolution, particularly for the onset of the event[17]. Indeed, it has been argued that the evidence for early warming is an artefact of sediment mixing or within the range of normal climatic variability[18,19]. To gain greater temporal resolution, we have generated chemical and biotic records from expanded PETM sediment sections located in shelf areas (Fig. 1). Two of these sections are from the New Jersey shelf at ~40 °N palaeolatitude: Ocean Drilling Program (ODP) Site 'Bass River' and United States Geological Survey (USGS) borehole 'Wilson Lake'. We generated carbon isotope and palynological data at these sites, in conjunction with new temperature records based on the organic palaeothermometer TEX_{86} ('tetraether index' of tetraether lipids consisting of 86 carbon atoms).

At Bass River, the onset of the CIE was previously identified by a negative step in the δ^{13}C of bulk carbonate ($\delta^{13}C_{BC}$) and foraminifera[20,21]. We have augmented these data by generating δ^{13}C records of dinocyst assemblages ($\delta^{13}C_{DINO}$) (Fig. 2a, Supplementary Fig. 1), and our resolution is over three times higher than the original study at this site[20]. The onset of the CIE is at 357.3 m below surface (m.b.s.) based on the $\delta^{13}C_{BC}$ record (Fig. 2a), the same level as indicated by δ^{13}C records from foraminifera[21]. Even though the upper part of the CIE is truncated by a sequence boundary at ~347.0 m.b.s. (Supplementary Fig. 1), the ~10.3-m-thick CIE interval is greatly expanded relative to that found in existing deep marine sections (<3 m). The onset of the *Apectodinium* acme is at 357.7 m.b.s., clearly lying below the base of the CIE by ~40 cm. This is not an artefact of bioturbation because the CIE is identified at the same level in the $\delta^{13}C_{DINO}$ record, which is predominantly derived from *Apectodinium* in this interval (Fig. 2a). Only the first ~120 kyr of the CIE is represented (Supplementary Information), so that sedimentation rates across this interval are 8–10 cm kyr^{-1}. Thus, the onset of the *Apectodinium* acme preceded the onset of the CIE at this site by about 4–5 kyr.

To assess the relative timing and magnitude of PETM warming at Bass River, we constructed a temperature record based on TEX_{86}. Planktonic foraminifer δ^{18}O records are not complete, exhibit much scatter and even within samples the variability in δ^{18}O of individual planktonic foraminifera is very large; hence, these records are not suitable for determining the onset of PETM warming (Supplementary Information). The TEX_{86} palaeothermometer is based on the

[1]Palaeoecology, Institute of Environmental Biology, Utrecht University, Laboratory of Palaeobotany and Palynology, [2]Department of Earth Sciences, Utrecht University, Budapestlaan 4, 3584 CD Utrecht, The Netherlands. [3]Royal Netherlands Institute for Sea Research (NIOZ), Department of Marine Biogeochemistry and Toxicology, PO Box 59, 1790 AB, Den Burg, Texel, The Netherlands. [4]Earth Sciences Department, University of California Santa Cruz, Santa Cruz, California 95060, USA. [5]Department of Earth Sciences, Rice University, 6100 Main Street, Houston, Texas 77005, USA. †Present addresses: Integrated Ocean Drilling Program, Texas A&M University, 1000 Discovery Drive, College Station, Texas 77845, USA (C.M.J.); GNS Science, PO Box 30-368, Lower Hutt, New Zealand (E.M.C.).

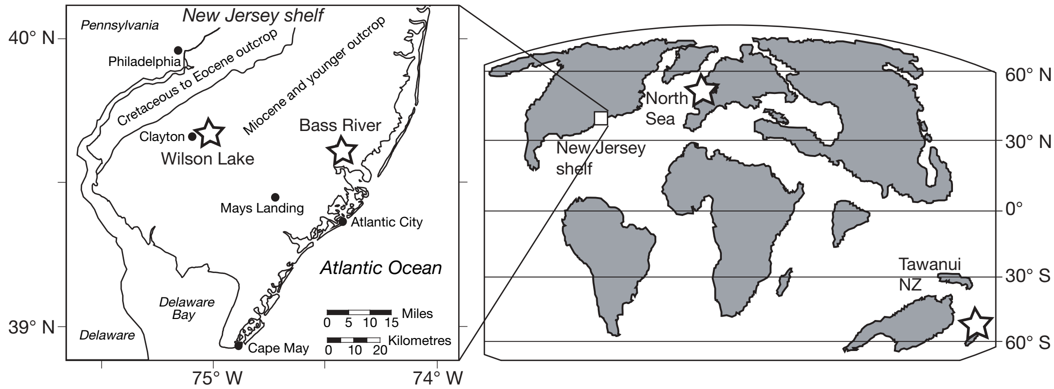


Figure 1 | Location of the studied sites. Left, drill locations of the Wilson Lake and Bass River cores. Right, locations of all studied sites within a palaeogeographic reconstruction of the Earth at PETM times (see Supplementary Information for the sources of the maps). NZ, New Zealand.

relative abundance of fossil crenarchaeotal membrane lipids in sediment. In modern marine surface sediments, TEX_{86} shows a significant linear correlation with mean annual sea surface temperature (SST), independent of nutrient availability or salinity[22,23]. The Bass River TEX_{86} values are up to 0.92, and thus exceed those measured in modern marine sediment (<0.8). Two temperature calibrations can be applied to these high TEX_{86} values, which suggest surface water warming during the PETM from 28 to 34 °C (ref. 24) or from 31 to 41 °C (ref. 22) (Supplementary Fig. 1). With either calibration, the TEX_{86} record at Bass River shows gradual warming between ~357.55 and 357.0 m.b.s., including intermediate temperatures (Fig. 2a). The record clearly indicates that onset of anomalous warming lies at least ~25–30 cm below the CIE, and that approximately half of the total warming occurred before the onset of the CIE (Fig. 2a). Assuming the above sedimentation rates, the onset of the *Apectodinium* acme started ~1–1.5 kyr before the onset of warming. In turn, the onset of warming at Bass River preceded the massive input of ^{13}C-depleted carbon by approximately 3–3.5 kyr.

To confirm the trends at Bass River, we generated a similar data set for the Wilson Lake site. At this site, the CIE is approximately 13.5 m thick, based on both the $\delta^{13}C_{BC}$ and the $\delta^{13}C_{DINO}$ record (Supplementary Fig. 2). In the $\delta^{13}C_{BC}$ record, the base of the CIE is not as clearly marked as at Bass River, primarily due to the absence of carbonate between ~109.3 and ~109.8 m.b.s. (Fig. 2b). Our $\delta^{13}C_{DINO}$ record shows background upper Palaeocene values, which are ~1‰ heavier than at Bass River, up to ~110.0 m.b.s. At this level the transition to true CIE values starts, coinciding with a negative step in the $\delta^{13}C_{BC}$ record, together implying the base of the CIE to be at this level. Several relatively heavy $\delta^{13}C_{DINO}$ values occur between ~110.6 and ~110.8 m.b.s., which are not present in the $\delta^{13}C_{BC}$

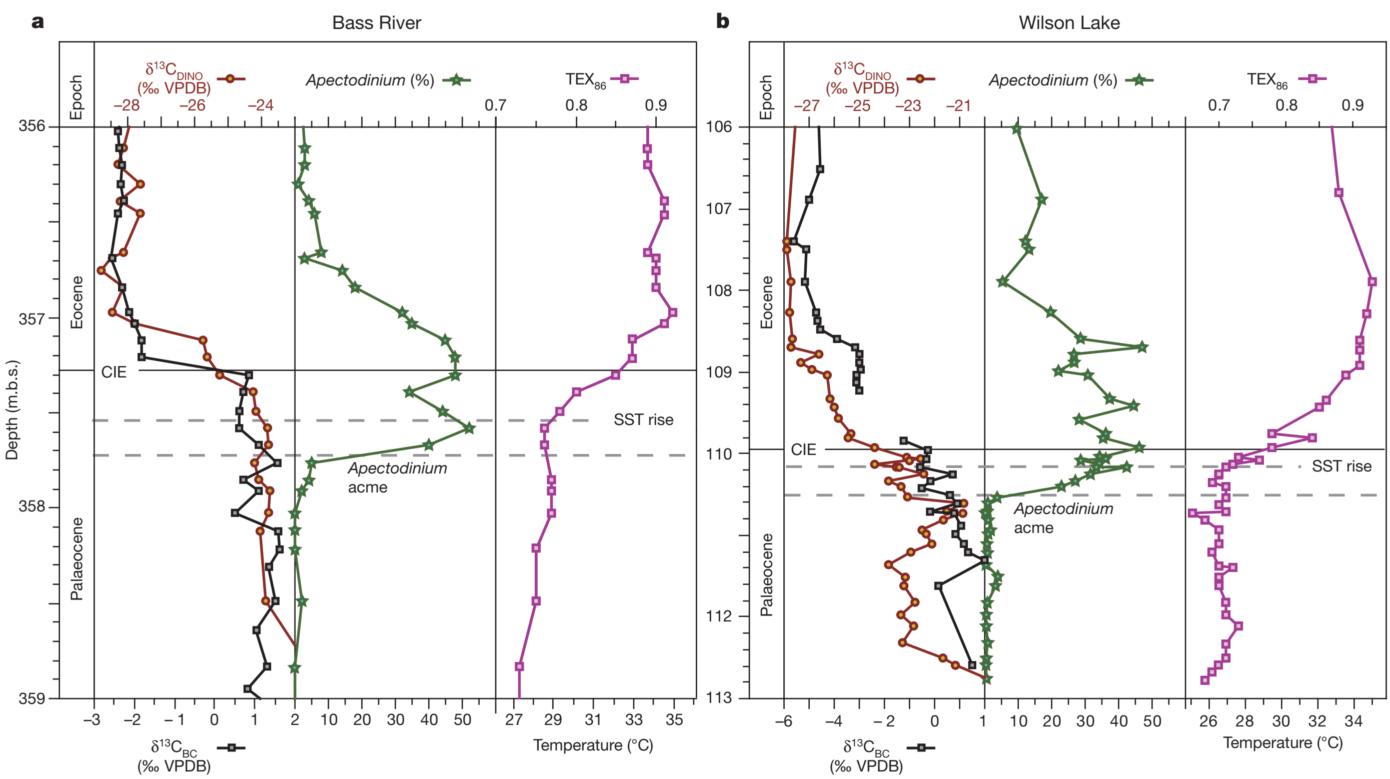


Figure 2 | High-resolution records across the onset of the PETM at the New Jersey Shelf Sites. **a**, Bass River. Bulk carbonate $\delta^{13}C$ data are from ref. 21. **b**, Wilson Lake. Part of the TEX_{86} data are from ref. 25 (see Supplementary Table 1 for details). Solid horizontal lines at ~357.3 m.b.s. and ~110.0 m.b.s. for Bass River and Wilson Lake, respectively, represent the onset of the CIE; dashed lines represent the onsets of the *Apectodinium* acme and surface warming. BC, bulk carbonate; DINO, dinocysts; VPDB, Vienna Pee Dee Belemnite; m.b.s., metres below surface. Scale at TEX_{86} temperatures represents the calibration from ref. 24. See Supplementary Information for the full records.

record. Such values are also not recorded in equivalent strata at Bass River (Fig. 2a), indicating they do not represent a global exogenic $\delta^{13}C$ change, and are likely to result from local factors influencing $\delta^{13}C_{DINO}$ at Wilson Lake. The onset of the *Apectodinium* acme at ~110.4 m.b.s. clearly lies below the CIE by ~40 cm. Given the 13.5 m thickness of the ~170 kyr CIE, sedimentation rates average ~8 cm kyr^{-1} at this site, so this offset represents about 4–5 kyr. Similar to Bass River, the TEX_{86} record at Wilson Lake, updated from ref. 25, indicates intermediate values and several data points suggest that warming occurred before the CIE (from ~110.15 m.b.s.). The pre-CIE warming is less clear at Wilson Lake than at Bass River, but the offset of ~15 cm implies that warming initiated ~2 kyr before the CIE (Fig. 2b).

The somewhat different records at the two sites might represent local processes that affect deposition in neritic settings. For example, Wilson Lake was located closer to shore (Fig. 1) and the record is likely to be less complete than that from Bass River because of breaks in deposition[20]. Although uncertainties in the age models necessarily imply some uncertainties in the duration of the leads and lags, the two New Jersey records show that a pulse in *Apectodinium* and subsequent sea surface water warming preceded the input of the ^{13}C-depleted carbon that caused the CIE by several thousand years.

Very few other locations are available where the sequence of events at the onset of PETM can be studied. However, the onset of the *Apectodinium* acme also appears to have started ~4 kyr before the CIE in a newly studied sediment section from the central North Sea at ~55 °N palaeolatitude, and, although less clearly, in published data from a section from the southwest Pacific Ocean at ~55 °S palaeolatitude (Supplementary Fig. 3; Supplementary Information). Hence, in mid-latitude regions, a marked increase in *Apectodinium* is the earliest sign of anomalous environmental change associated with the PETM. A baseline requirement for abundant *Apectodinium* appears to have been relatively high temperatures, judging from biogeographical patterns[13]. As *Apectodinium* was abundant in the Arctic Ocean when SSTs reached 23 °C (ref. 6), it should have flourished on the New Jersey shelf during the late Palaeocene if SST was the only environmental control on its abundance. Dinoflagellate assemblages are extremely sensitive to their surroundings, and significant changes in other environmental parameters, perhaps including surface water stratification, salinity and food supply, must have contributed to the *Apectodinium* acme at middle to high latitudes during the PETM[8,13] (Supplementary Information).

One possible constraint on their environmental preferences is derived from recent data from New Jersey (Supplementary Fig. 4) and the Arctic[6]. These indicate that *Apectodinium* locally became outnumbered by typical low-salinity-tolerant taxa, which suggests that very low salinities were not optimal for *Apectodinium*. Occasional pulses of *Apectodinium* in upper Palaeocene strata deposited in the Tethyan Ocean indicate that appropriate conditions occurred episodically and locally at low latitudes before the PETM (Supplementary Information). Such unusual conditions, however, only became globally and uniquely significant at the onset of the PETM. Although surface water parameters causing the *Apectodinium* acme remain unknown, the acme serves as a harbinger to the extensive global environmental change at the PETM.

Warming before the CIE is consistent with the hypothesis that thermal dissociation of submarine gas hydrates caused the CIE[1]. In the present oceans, large amounts of gas hydrate exist within sediments along continental slopes where microbes generate sufficient methane to saturate pore waters at relatively high pressures and low temperatures[26]. Although the distribution of gas hydrates was probably less extensive in the Palaeogene because of warmer deep waters, the mass of gas hydrates may have been similar to or greater than present-day amounts because of enhanced methane production[27]. Given mass estimates for modern gas hydrates[28], a rapid 5 °C increase in deep ocean temperature, such as inferred for the PETM[3,5,14], should have released nominally 2,000 Gt of methane, about that required to cause a –3‰ excursion in global $\delta^{13}C$ records[1]. Crucially, this mechanism necessitates a time lag between initial warming and carbon input because heat must propagate downward from the sea floor to drive an endothermic reaction in underlying sediment. The duration of this time lag should be several thousands of years[9,10], which agrees with observations in our records.

Our records clearly show that initiation of the *Apectodinium* acme and substantial, rapid warming along the ancient shelf of New Jersey preceded the massive input of ^{13}C-depleted carbon that caused the CIE. This observation provides a fundamental constraint on mechanisms for global change during the PETM. Potentially, the pre-CIE warming in New Jersey shelf waters was caused by substantial changes in ocean–atmosphere circulation—perhaps associated with an orbitally modulated[29] surpassing of a threshold—which was the initially proposed mechanism for methane hydrate dissociation at the PETM[1]. Alternatively, the warming was driven by an increase in greenhouse gas concentrations that did not significantly affect the $\delta^{13}C$ of the exogenic carbon pool and, hence, our $\delta^{13}C$ records. An enormous input of mantle-derived carbon is possible because evidence exists for massive volcanism close to the PETM[30], and because such an injection (with a $\delta^{13}C$ of –5‰ to –7‰) would have minimally affected the $\delta^{13}C$ of the ocean–atmosphere reservoir. However, it remains uncertain whether volcanism could have produced CO_2 at a rate required for rapid global environmental change. Higher net fluxes of CO_2 from the ocean into the atmosphere could also have caused higher atmospheric concentrations without significantly changing global exogenic $\delta^{13}C$. Pre-CIE warming also could have been caused by an increase in non-carbon greenhouse gases. If the pre-CIE warming was carbon-forced, it directly implies that the warming and ocean acidification during the PETM were caused by at least two sources of carbon with different isotopic compositions, compromising the use of simple mass balance calculations[2,12] to elucidate the origin and volume of the injected carbon and the estimation of climate sensitivity to carbon input. It is important to note that although the *Apectodinium* acme started before the CIE, the dinoflagellate assemblages did not change significantly across the onset of the CIE. This suggests that environmental change was well under way before the injection of ^{13}C-depleted carbon, at least in New Jersey and the North Sea.

METHODS SUMMARY

Freeze-dried samples were processed for fully quantitative analyses (including addition of the exote *Lycopodium*) using 30% HCl and 38% HF. Processing of the North Sea samples (Supplementary Information) included oxidation in Schultz's reagent at 80 °C. Residues were placed in an ultrasonic bath for a maximum of 5 min and sieved over a 15-μm mesh. Slides were analysed at 500× magnification to a minimum of 200 dinocysts. Absolute quantitative numbers were calculated using the relative number of *Lycopodium*.

Extraction of crenarchaeotal tetraether lipids was performed on freeze-dried samples using accelerated solvent extraction with a dichloromethane/methanol (2:1) mixture, after which polar and apolar fractions were separated. Polar fractions were filtered and analysed using high performance liquid chromatography/atmospheric pressure chemical ionization-mass spectrometry. Single ion monitoring was used to quantify the abundance of the crenarchaeotal lipids. TEX_{86} values were determined at least in duplicate and converted to temperature using the equations from refs 22 and 24 (see text). Abundances of branched tetraether lipids from terrestrial origin appeared very low. The crenarchaeotal lipids are not present in the sediments from the North Sea site due to the relatively high maturity of the organic matter.

For the $\delta^{13}C_{DINO}$ records, dinocysts were isolated using a 40 μm and a 125 μm nylon mesh sieve and oven-dried. All $\delta^{13}C_{TOC}$ (North Sea site; Supplementary Information) and $\delta^{13}C_{DINO}$ analyses were done with a Fison NA 1500 CNS analyser, connected to a Finnigan Delta Plus mass spectrometer. Analytical precision determined by replicate analyses was better than 0.1‰. For the $\delta^{13}C_{BC}$ values, freeze-dried samples were measured using an Autocarb or Optima coupled to a PRISM Mass Spectrometer at UCSC. Precision, based on replicate analyses of in-house standard Carrara Marble, is better than ±0.05‰. All isotope values are reported relative to the Vienna Pee Dee Belemnite (VPDB) standard.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 June; accepted 18 October 2007.

1. Dickens, G. R., O'Neil, J. R., Rea, D. K. & Owen, R. M. Dissociation of oceanic methane hydrate as a cause of the carbon isotope excursion at the end of the Paleocene. *Paleoceanography* **10,** 965–971 (1995).
2. Pagani, M., Caldeira, K., Archer, D. & Zachos, J. C. An ancient carbon mystery. *Science* **314,** 1556–1557 (2006).
3. Kennett, J. P. & Stott, L. D. Abrupt deep-sea warming, palaeoceanographic changes and benthic extinctions at the end of the Palaeocene. *Nature* **353,** 225–229 (1991).
4. Zachos, J. C. *et al.* A transient rise in tropical sea surface temperature during the Paleocene-Eocene thermal maximum. *Science* **302,** 1551–1554 (2003).
5. Tripati, A. & Elderfield, H. Deep-sea temperature and circulation changes at the Paleocene-Eocene thermal maximum. *Science* **308,** 1894–1898 (2005).
6. Sluijs, A. *et al.* Subtropical Arctic Ocean temperatures during the Palaeocene/Eocene thermal maximum. *Nature* **441,** 610–613 (2006).
7. Sluijs, A., Bowen, G. J., Brinkhuis, H., Lourens, L. J. & Thomas, E. in *Deep Time Perspectives on Climate Change: Marrying the Signal from Computer Models and Biological Proxies* (eds Williams, M., Haywood, A. M., Gregory, F. J. & Schmidt, D. N.) 323–349 (The Micropalaeontological Society Special Publication, The Geological Society, London, 2007).
8. Crouch, E. M. *et al.* Global dinoflagellate event associated with the late Paleocene thermal maximum. *Geology* **29,** 315–318 (2001).
9. MacDonald, G. J. Role of methane clathrates in past and future climates. *Clim. Change* **16,** 247–281 (1990).
10. Xu, W., Lowell, R. P. & Peltzer, E. T. Effect of seafloor temperature and pressure variations on methane flux from a gas hydrate layer: Comparison between current and late Paleocene climate conditions. *J. Geophys. Res.* **106,** 26413–26423 (2001).
11. Koch, P. L., Zachos, J. C. & Gingerich, P. D. Correlation between isotope records in marine and continental carbon reservoirs near the Palaeocene/Eocene boundary. *Nature* **358,** 319–322 (1992).
12. Higgins, J. A. & Schrag, D. P. Beyond methane: Towards a theory for the Paleocene-Eocene thermal maximum. *Earth Planet. Sci. Lett.* **245,** 523–537 (2006).
13. Bujak, J. P. & Brinkhuis, H. in *Late Paleocene-Early Eocene Climatic and Biotic Events in the Marine and Terrestrial Records* (eds Aubry, M.-P., Lucas, S. G. & Berggren, W. A.) 277–295 (Columbia Univ. Press, New York, 1998).
14. Thomas, E. & Shackleton, N. J. in *Correlation of the Early Paleogene in Northwestern Europe* (eds Knox, R. W. O. B., Corfield, R. M. & Dunay, R. E.) 401–441 (Geological Society of London Special Publication 101, Geological Society of London, London, 1996).
15. Svensen, H. *et al.* Release of methane from a volcanic basin as a mechanism for initial Eocene global warming. *Nature* **429,** 542–545 (2004).
16. Thomas, D. J., Zachos, J. C., Bralower, T. J., Thomas, E. & Bohaty, S. Warming the fuel for the fire: Evidence for the thermal dissociation of methane hydrate during the Paleocene-Eocene thermal maximum. *Geology* **30,** 1067–1070 (2002).
17. Zachos, J. C. *et al.* Rapid acidification of the ocean during the Paleocene-Eocene thermal maximum. *Science* **308,** 1611–1615 (2005).
18. Cramer, B. S. & Kent, D. V. Bolide summer: The Paleocene/Eocene thermal maximum as a response to an extraterrestrial trigger. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **224,** 144–166 (2005).
19. Stoll, H. M. Limited range of interspecific vital effects in coccolith stable isotopic records during the Paleocene-Eocene thermal maximum. *Paleoceanography* **20,** doi:10.1029/2004PA001046 (2005).
20. Cramer, B. S. *et al.* An exceptional chronologic, isotopic, and clay mineralogic record of the latest Paleocene thermal maximum, Bass River, NJ, ODP 174AX. *Bull. Soc. Geol. Fr.* **170,** 883–897 (1999).
21. Zachos, J. C. *et al.* The Palaeocene-Eocene carbon isotope excursion: Constraints from individual shell planktonic foraminifer records. *Phil. Trans. R. Soc. A* **365,** 1829–1842 (2007).
22. Schouten, S., Hopmans, E. C., Schefuß, E. & Sinninghe Damsté, J. S. Distributional variations in marine crenarchaeotal membrane lipids: A new tool for reconstructing ancient sea water temperatures? *Earth Planet. Sci. Lett.* **204,** 265–274 (2002).
23. Wuchter, C., Schouten, S., Coolen, M. J. L. & Sinninghe Damsté, J. S. Temperature-dependent variation in the distribution of tetraether membrane lipids of marine Crenarchaeota: Implications for TEX_{86} paleothermometry. *Paleoceanography* **19,** doi:10.1029/2004PA001041 (2004).
24. Schouten, S. *et al.* Extremely high sea-surface temperatures at low latitudes during the middle Cretaceous as revealed by archaeal membrane lipids. *Geology* **31,** 1069–1072 (2003).
25. Zachos, J. C. *et al.* Extreme warming of mid-latitude coastal ocean during the Paleocene-Eocene thermal maximum: Inferences from TEX_{86} and isotope data. *Geology* **34,** 737–740 (2006).
26. Kvenvolden, K. A. Methane hydrate — a major reservoir of carbon in the shallow geosphere? *Chem. Geol.* **71,** 41–51 (1988).
27. Dickens, G. R. Rethinking the global carbon cycle with a large, dynamic and microbially mediated gas hydrate capacitor. *Earth Planet. Sci. Lett.* **213,** 169–183 (2003).
28. Milkov, A. V. Global estimates of hydrate-bound gas in marine sediments: How much is really out there? *Earth Sci. Rev.* **66,** 183–197 (2004).
29. Lourens, L. J. *et al.* Astronomical pacing of late Palaeocene to early Eocene global warming events. *Nature* **435,** 1083–1087 (2005).
30. Storey, M., Duncan, R. A. & Swisher, C. C. III. Paleocene-Eocene thermal maximum and the opening of the northeast Atlantic. *Science* **316,** 587–589 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This research used samples and/or data provided by the Integrated Ocean Drilling Program. A.S. thanks the Utrecht Biogeology Center, the LPP Foundation and The Netherlands Organisation for Scientific Research (NWO, VENI-grant) and J.C.Z. thanks the National Science Foundation for funding. E. Hopmans and A. Mets (NIOZ) are thanked for analytical assistance with TEX_{86} measurements. A. van Dijk, E. van Bentum, C. Blaga, R. Deltrap, D. Menzel, G. Nobbe, J. van Tongeren, N. Welters (all Utrecht University) and D. Andreason and T. Quattlebaum (UC Santa Cruz) are thanked for advice and support regarding analyses. We thank J. Browning, K. Miller, J. Firth and G. Esmay for help regarding the logistics around sampling the Bass River section and E. Thomas for sampling the Wilson Lake section. ConocoPhillips, P. van Veen in particular, is thanked for enabling us to use their North Sea cores. I. Harding, M. Huber, L. Lourens and M. Pagani are thanked for discussions.

Author Contributions A.S., H.B. and J.C.Z. designed the research, A.S. and H.B. carried out the palynology, A.S., G.-J.R. and H.B. the $\delta^{13}C_{DINO}$ and $\delta^{13}C_{TOC}$ analyses, A.S., S.S., G.J.R. and J.S.S.D. the TEX_{86} analyses. S.M.B., C.M.J. and J.C.Z. generated the isotope data on carbonate. All authors contributed to interpreting the data and writing the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.S. (A.Sluijs@uu.nl).

doi:10.1038/nature06400

METHODS

Palynology. To ~10 g of freeze-dried sediment, a known amount of *Lycopodium* spores was added. Samples were treated with 30% HCl and twice with 38% HF for carbonate and silicate removal, respectively, and sieved over a 15-μm nylon mesh. Samples from the North Sea (Supplementary Information) were sieved in a 15-μm steel mesh sieve and subsequently oxidized for 20 min in Schultz's reagent at 80 °C to reduce the amount of amorphous organic matter. To disaggregate clumps of residue, samples were placed in an ultrasonic bath for a maximum of 5 min, sieved again, and subsequently concentrated to 1 ml, of which 7.5–10 μl was mounted on microscope slides. Slides were analysed at 500× magnification to a minimum of 200 dinocysts. Absolute quantitative numbers were calculated using the relative number of *Lycopodium*.

TEX_{86} analyses. Powdered and freeze-dried sediments (~20 g dry mass) were extracted with a Dionex Accelerated Solvent Extractor using a 9:1 (v:v) mixture of dichloromethane (DCM) and methanol (MeOH). The extract was fractionated into apolar and polar fractions, containing the crenarchaeotal lipids, using a small column with activated alumina and using hexane/DCM (9:1;v/v) and DCM/MeOH (1:1;v/v) as eluents, respectively. Polar fractions were dissolved in hexane/propanol (99:1;v/v), and filtered through 0.45 μm PTFE filters. The samples were analysed with an Agilent 1100 series LC/MSD SL and separation was performed on an Prevail Cyano column (2.1 × 150 mm, 3 μm; Alltech), maintained at 30 °C. The glycerol dialkyl glycerol tetraethers (GDGTs) were eluted using a changing mixture of hexane and propanol as follows, 99 hexane:1 propanol for 5 min, then a linear gradient to 1.8 propanol in 45 min. Flow rate was 0.2 ml min^{-1}. Detection was achieved using atmospheric pressure chemical ionization-mass spectrometry of the eluent. Single ion monitoring was set to scan the 5 $[M+H]^+$ ions of the GDGTs with a dwell time of 237 ms for each ion. TEX_{86} values were determined at least in duplicate and converted to temperature using the equations from refs 22 and 24 (see text). In addition, we measured the abundance of branched and isoprenoid tetraether lipids to estimate the relative abundance of terrestrially derived GDGTs, which appeared very low. GDGTs are not present in the sediments from the North Sea site owing to the relatively high maturity of the organic matter.

Organic $\delta^{13}C$ measurements. For the $\delta^{13}C_{DINO}$ records, the 40–125 μm size fraction of the palynological residues, which are nearly barren of organic particles other than dinocysts, were isolated using nylon mesh sieves, and oven-dried. For $\delta^{13}C_{TOC}$ analyses (North Sea site; Supplementary Information), samples were freeze-dried and powdered. All $\delta^{13}C$ analyses were done with a Fison NA 1500 CNS analyser, connected to a Finnigan Delta Plus mass spectrometer. Analytical precision determined by replicate analyses was better than 0.1‰. All values are reported relative to the VPDB standard.

Bulk carbonate $\delta^{13}C$ measurements. Freeze-dried samples were measured using an Autocarb or Optima coupled to a PRISM Mass Spectrometer at UCSC. Precision based on replicate analyses of in-house standard Carrara Marble is better than ±0.05‰. All values are reported relative to the VPDB standard.

LETTERS

Trench-parallel flow and seismic anisotropy in the Mariana and Andean subduction systems

Erik A. Kneller[1] & Peter E. van Keken[1]

Shear-wave splitting measurements above the mantle wedge of the Mariana[1] and southern Andean[2,3] subduction zones show trench-parallel seismically fast directions close to the trench and abrupt rotations to trench-perpendicular anisotropy in the back arc. These patterns of seismic anisotropy may be caused by three-dimensional flow associated with along-strike variations in slab geometry[1–5]. The Mariana and Andean subduction systems are associated with the largest along-strike variations of slab geometry observed on Earth[6,7] and are ideal for testing the link between slab geometry and solid-state creep processes in the mantle. Here we show, with fully three-dimensional non-newtonian subduction zone models, that the strong curvature of the Mariana slab and the transition to shallow slab dip in the Southern Andes give rise to strong trench-parallel stretching in the warm-arc and warm-back-arc mantle and to abrupt rotations in stretching directions that are accompanied by strong trench-parallel stretching. These models show that the patterns of shear-wave splitting observed in the Mariana and southern Andean systems may be caused by significant three-dimensional flow induced by along-strike variations in slab geometry.

Shear-wave splitting occurs when shear waves propagate through anisotropic material and split into two orthogonally polarized waves that travel at different velocities[5]. This type of seismic anisotropy is a straightforward indicator of elastic structure and provides information about lattice preferred orientation and flow geometry within the mantle[5]. The most commonly observed fabric in dry peridotite is A-type, which causes alignment of the seismically fast splitting axis with the maximum stretch direction or flow direction for simple shear. Shear-wave splitting observations from most subduction zones show complex patterns of seismic anisotropy that usually have trench-parallel fast directions[8]. These observations are unexpected, because models of two-dimensional wedge flow predict fast seismic anisotropy parallel to plate motion (trench perpendicular)[9]. Possible causes for trench-parallel anisotropy and abrupt rotations in fast directions in the mantle wedge of subduction zones include olivine fabric transitions[10–12], melt-related anisotropy[13,14] and three-dimensional flow with stretching-parallel olivine fabrics[4]. Three-dimensional flow may be caused by small-scale convection[15], oblique subduction[16], differential slab rollback[17], trench-parallel motion of the overriding plate[4], and variations in slab geometry[4,16]. Testing the applicability of these models to specific subduction systems will improve our ability to infer flow geometry from seismic observations and provide better constraints on thermal structure and geochemical transport processes.

Whereas melt-related processes and olivine fabric transitions are applicable to local regions of the mantle wedge[8,12], three-dimensional flow has the potential to explain the trench-parallel shear-wave splitting that is widespread in some subduction zones. Here we explore the effects of slab geometry on deformation in the mantle wedge of segments of the Mariana and Andean subduction systems. The modelled segments show strong along-strike variation in slab curvature (Mariana) and dip (Andes). Both geometrical changes induce along-strike pressure changes that lead to significant three-dimensional flow. Accurate modelling of the magnitude of the velocity and strain geometry in the mantle wedge requires fully three-dimensional high-resolution calculations with realistic stress- and temperature-dependent rheology (see Methods Summary).

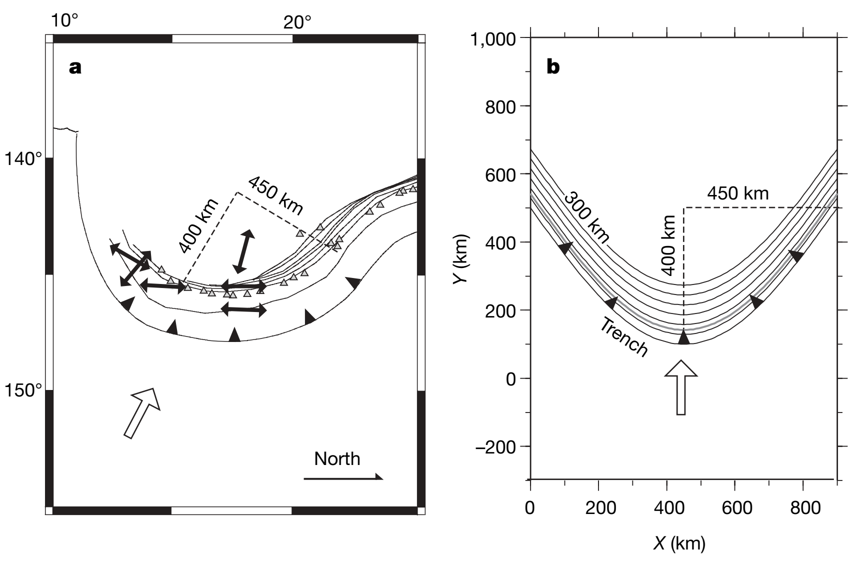


Figure 1 | Observations and approximate slab model of the Mariana subduction system. Contours show the depth to slab at 50 km intervals[6]. The hachured lines (solid black triangles) indicate the trench and dip direction. Dashed lines show the length scales of slab curvature. The open arrow denotes the relative convergence direction of the Pacific plate with respect to the Philippine plate. **a**, Geometry of the Mariana subduction system between 10°–25° N (horizontal axis) and 135°–155° E (vertical axis). Black double arrows show the pattern of seismically fast directions associated with local-shear-wave phases[1]. The positions of the Holocene volcanoes (grey triangles) are from the Smithsonian Institution, Global Volcanism Program (http://www.volcano.si.edu). **b**, Simplified slab model. Black solid lines denote 50-km slab contours, whereas the grey line denotes the depth at which full viscous coupling between the slab and mantle wedge starts.

[1]Department of Geological Sciences, University of Michigan, Ann Arbor, Michigan 48109, USA.

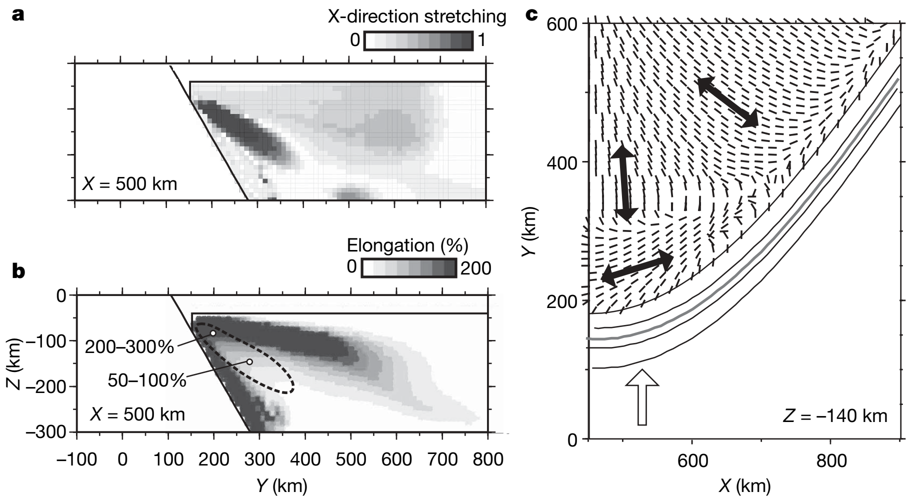


Figure 2 | Finite strain calculations from the Mariana model. **a**, **b**, Averaged finite strain calculations along vertical cross-sections. **c**, A horizontal cross-section with the maximum stretch directions plotted as thin short lines. The stretching magnitude in the *X*-direction is shown in **a** (colour scale) and is equal to the absolute value of the convergence normal component of the maximum stretch axis over the magnitude of maximum stretch. When this quantity is equal to one, the maximum stretch direction is in the horizontal plane and parallel to the *X*-axis (approximately convergence-normal). The maximum elongation (colour scale) is shown in **b**, where the region of maximum trench-parallel stretching from **a** is indicated by the black dashed ellipse. The heavy black arrows indicate the splitting pattern that would follow from this geometry of finite strain. The grey contour in **c** denotes the depth of full mechanical coupling between the slab and mantle wedge.

The characteristic amplitude and width of the curvature of the Mariana slab are approximately 400 and 900 km, respectively[6] (Fig. 1a). We construct a simplified model of the Mariana slab with a constant slab dip of 60° in the convergence direction and describe slab curvature with a harmonic function (Fig. 1b; Supplementary Fig. 1b). The magnitude of the trench-parallel velocity in the mantle wedge increases towards the trench and reaches a maximum in the arc mantle (Supplementary Fig. 2a and b). This is caused by the strong curvature in the slab and gives rise to a radiating flow geometry. Strong focusing of the flow occurs where curvature is largest, owing to the local decrease in viscosity associated with the strain-rate-dependent and temperature-dependent rheology.

Finite-strain calculations for the Mariana model show trench-parallel stretching in the arc mantle and an abrupt change to trench-perpendicular stretching in the back arc (Fig. 2). This region of strong trench-parallel stretching is associated with the strongest velocity gradients and is located between 80 and 150 km depth (Fig. 2a and b). The maximum thickness is around 50 km, which is consistent (assuming maximum fabric strength and a total intrinsic anisotropy of around 6–7%; ref. 18) with a maximum delay time for split shear waves of approximately 0.5–1 s. Recent local-shear-wave splitting measurements above the mantle wedge of the Mariana subduction zone show trench-parallel fast directions in the arc mantle, a rotation to trench perpendicular in the back arc, and delay times of around 1 s (ref. 1). The similarities between shear-wave splitting observations (Fig. 1a) and calculated stretching patterns (Fig. 2c) suggest that anisotropy in the Mariana wedge is controlled by three-dimensional stretching induced by slab curvature.

In the Andean subduction system large variations in slab dip occur near flat slab subduction[6,7]. For example, a transition from average slab dip equal to 30° to less than 5° over 400 km is observed adjacent to the flat slab in the Chile–Argentina region[2,3,7] (Fig. 3a). We construct an approximate model of the Chile–Argentina slab, using a linear trench and a harmonic function that describe slab dip variations along strike (Fig. 3b; Supplementary Fig. 1c). For the Andean model, average slab dip varies from 30° at $X = 400$ km to 10° at $X = 0$ km, which is similar to the geometry of the Chile–Argentina slab (Supplementary Fig. 3).

Our fluid dynamic model shows relatively low dynamic pressure near the shallow dipping slab, which drives strong trench-parallel flow (Supplementary Fig. 4). The calculated flow geometry consists of flow lines that are parallel to the trench near the shallow dipping slab and trench-perpendicular directly above the shallow dipping slab. The magnitude of trench-parallel flow reaches a maximum close to the convergence velocity in the region where the slab is at 75–100 km depth (Supplementary Fig. 4c and d). Strong trench-parallel focusing of flow occurs in the corner of the mantle wedge next to the shallow dipping slab (Supplementary Fig. 4a). This flow geometry leads to a complicated pattern of deformation with strong trench-parallel stretching in the shallow mantle to the right (south) of the shallow dipping segment. The stretching abruptly becomes trench-perpendicular directly above the shallow dipping slab (Fig. 4a and b).

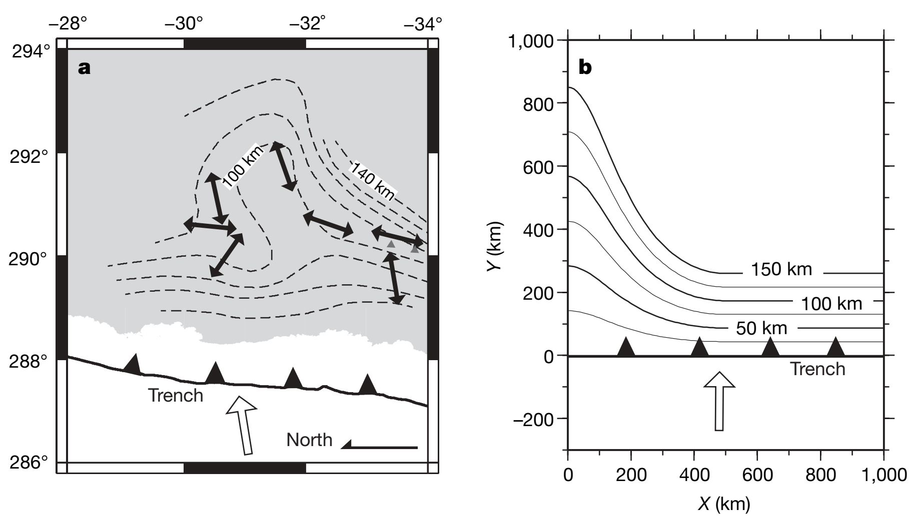


Figure 3 | Observations and approximate slab model of the Andean subduction system. **a**, Dashed lines show contours (10 km contour interval) of the subducting slab in the Northern Chile region of the Andean subduction system[7]. The grey shading in **a** indicates land. **b**, Simplified slab geometry used in this study. The average slab dip in **a** varies from less than 5° to 30° over 400 km. Bold arrows show the observed fast splitting directions[2,3]. The slab geometry used in the Andean model presented in this work (**b**) has a similar variation in average slab dip (10° to 30°) (Supplementary Fig. 1c). See Fig. 1 for further details.

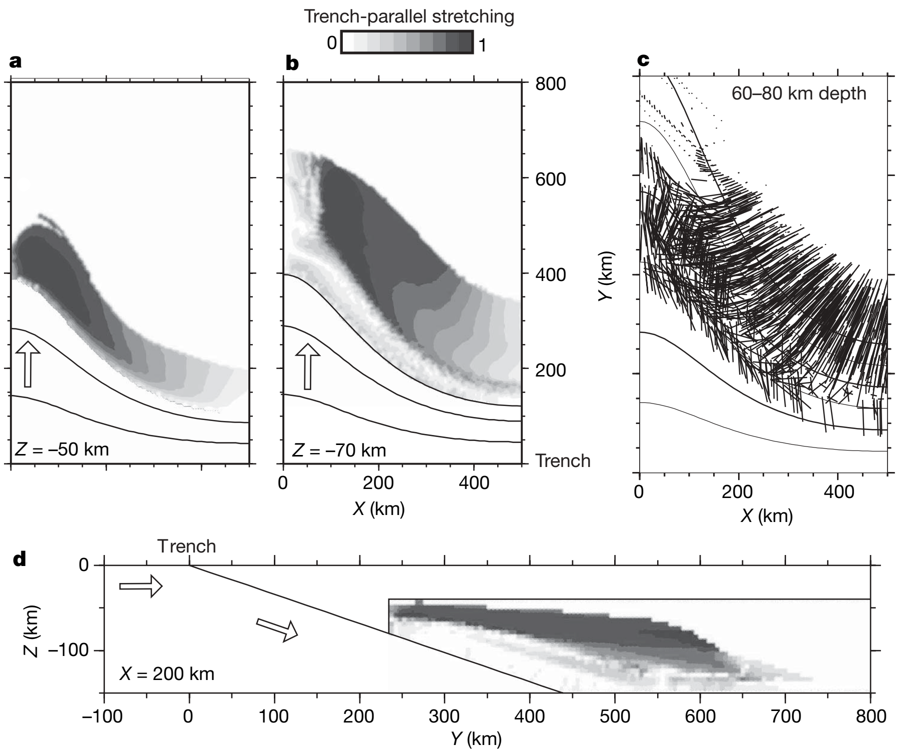


Figure 4 | Finite strain calculations from the Andean model. **a**, **b**, Maps showing the magnitude of trench-parallel stretching. **c**, Maximum stretch axes. **d**, Vertical cross-section of trench-parallel stretching at $X = 200$ km for the Andean model. See Fig. 2 for details. The magnitude of trench-parallel stretching is shown only for particles that have undergone greater than 50% elongation. This slab geometry is associated with thick layers of material with convergence-normal stretching in the arc and back-arc mantle.

A relatively thick layer of material with trench-parallel stretching develops in the mantle wedge with strain magnitudes exceeding 300% (Fig. 4c).

Trench-parallel shear-wave splitting and a rotation to trench-perpendicular splitting are observed with both teleseismic and local-shear-wave phases in the Chile–Argentina region of the Andean system[2,3] (Fig. 3a). These observations suggest the presence of strong trench-parallel stretching in the arc mantle, strong focusing and stretching in the corner of the mantle wedge adjacent to the flat slab, and a rapid rotation to trench-perpendicular stretching towards the flat slab segment. This inferred pattern of deformation (Fig. 3a) is strikingly similar to the pattern of stretching from the model presented here (Fig. 4a and b). This suggests that the strain field within this portion of the Andean wedge is controlled by three-dimensional variations in slab dip. Slab dip variations may also play an important role in controlling seismic anisotropy in the mantle wedge next to the flat slab of the Northern Andean Subduction system[19], which is associated with a similar transition to shallow dip, but over a distance of 900–1,000 km.

Three-dimensional flow plays an important role in a variety of processes occurring in the subduction factory such as thermal structure, geochemical transport, and rock fabric development. The shear-wave splitting observations and the calculations presented in this work demonstrate that models of slab metamorphism, arc magma production, and rock fabric development should consider three-dimensional flow in systems with large variations in slab geometry. We conclude that strong three-dimensional flow exists in the Mariana and Andean subduction systems and that this flow significantly affects the formation of seismic anisotropy in the mantle wedge.

METHODS SUMMARY

Solid-state creep in the mantle wedge is driven by a kinematically prescribed slab with along-strike variations of slab dip and trench-curvature defined by simple functions[20] (Supplementary Fig. 1). An experimentally based temperature- and stress-dependent rheology is implemented in the viscous wedge[21,22]. Temperature, velocity and dynamic pressure are obtained by numerically solving equations that govern heat transport and fluid flow with the finite element method. Large gradients in velocity and viscosity require the implementation of a high-resolution mesh with strong local refinement in the corner of the mantle wedge. Particle motion and finite strain are calculated using the steady-state velocity field. Integrated maximum stretch directions are used as proxy for olivine lattice preferred orientation and seismically fast directions in the warm-arc and warm-back-arc mantle. This proxy provides a reasonable approximation for olivine lattice-preferred-orientation development throughout most of the mantle wedge for the cases considered in this study because particles experience consistent velocity gradients as they travel through the mantle wedge[23,24].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 May; accepted 18 October 2007.

1. Pozgay, S. H., Wiens, D. A., Conder, J. A., Shiobara, H. & Sugioka, H. Complex mantle flow in the Mariana subduction system: evidence from shear wave splitting. *Geophys. J. Int.* 170, 371–386 (2007).
2. Anderson, M. L., Zandt, G., Triep, E., Fouch, M. & Beck, S. Anisotropy and mantle flow in the Chile-Argentina subduction zone from shear wave splitting analysis. *Geophys. Res. Lett.* 31, doi:10.1029/2004GL020906 (2004).
3. Anderson, M., Zandt, G. & Wagner, L. Along-strike mantle flow variations in a segment of the South American Subduction zone, Chile and Argentina. *Earth Planet. Sci. Lett.* (submitted).
4. Hall, C. E., Fischer, K. M., Parmentier, E. M. & Blackman, D. K. The influence of plate motions on three-dimensional back arc mantle flow and shear wave splitting. *J. Geophys. Res.* 105, 28009–28033 (2000).
5. Park, J. & Levin, V. Seismic anisotropy: tracing plate dynamics in the mantle. *Science* 296, 485–489 (2002).
6. Syracuse, E. M. & Abers, G. A. Global compilation of variations in slab depth beneath arc volcanoes and implications. *Geochem. Geophys. Geosyst.* 7, doi:10.1029/2005GC001045 (2006).
7. Anderson, M., Alvarado, P., Zandt, G. & Beck, S. Geometry and brittle deformation of the subducting Nazca plate, central Chile and Argentina. *Geophys. J. Int.* 171, doi:10.1111/j.1365–246X.2007.03483.x (2007).
8. Wiens, D. A. & Smith, G. P. in *Inside The Subduction Factory* (ed. Eiler, J.) 83–105 (Geophysical Monograph 138, American Geophysical Union, Washington DC (2003).
9. Fischer, K. M., Parmentier, E. M., Stine, A. R. & Wolf, E. Modelling anisotropy and plate-driven flow in the Tonga subduction zone back arc. *J. Geophys. Res.* 105, 181–191 (2000).
10. Kneller, E. A., van Keken, P. E., Karato, S. & Park, J. B-type olivine fabric in the mantle wedge: Insights from high-resolution non-Newtonian subduction zone models. *Earth Planet. Sci. Lett.* 237, 781–797 (2005).
11. Katayama, I. & Karato, S. Effect of temperature on the B- to C-type olivine fabric transition and implication for flow pattern in subduction zones. *Phys. Earth Planet. Inter.* 157, 33–45 (2006).

12. Kneller, E. A., van Keken, P. E., Katayama, I. & Karato, S. Stress, strain, and B-type olivine fabric in the fore-arc mantle: sensitivity tests using high-resolution steady-state subduction zone models. *J. Geophys. Res.* **112,** B04406, doi:10.1029/2006JB004544 (2007).
13. Holtzman, B. K. *et al.* Melt segregation and strain partitioning: implications for seismic anisotropy and mantle flow. *Science* **301,** 1227–1230 (2003).
14. Kendall, J., Stuart, G. W., Ebinger, C. J., Bastow, I. D. & Keir, D. Magma-assisted rifting in Ethiopia. *Nature* **433,** 146–148 (2005).
15. Honda, S. & Saito, M. Small-scale convection under the back-arc occurring in the low viscosity wedge. *Earth Planet. Sci. Lett.* **216,** 703–715 (2003).
16. Mehl, L., Hacker, B. R., Hirth, G. & Kelemen, P. B. Arc-parallel flow within the mantle wedge. Evidence from the accreted Talkeetna arc, south central Alaska. *J. Geophys. Res.* **108,** doi:10.1029/2002JB002233 (2003).
17. Russo, R. M. & Silver, P. G. Trench-parallel flow beneath the Nazca plate from seismic anisotropy. *Science* **263,** 1105–1111 (1994).
18. Ismaïl, B. W. & Mainprice, D. An olivine fabric database: an overview of upper mantle fabrics and seismic anisotropy. *Tectonophysics* **196,** 145–157 (1998).
19. Kaneshima, S. & Silver, P. G. Anisotropic loci in the mantle beneath central Peru. *Phys. Earth Planet. Inter.* **88,** 257–272 (1995).
20. Kneller, E. A. & van Keken, P. E. The effect of three-dimensional slab geometry on deformation in the mantle wedge: implications for shear wave anisotropy. *Geochem. Geophys. Geosyst.* (in the press).
21. Karato, S. & Jung, H. Effects of pressure on high-temperature dislocation creep in olivine. *Phil. Mag.* **83,** 401–414 (2003).
22. Mei, S. & Kohlstedt, D. L. Influence of water on plastic deformation of olivine aggregates. 1. Diffusion creep regime. *J. Geophys. Res.* **105,** 21457–21469 (2000).
23. Kaminski, E. & Ribe, N. M. Time scales for the evolution of seismic anisotropy in mantle flow. *Geochem. Geophys. Geosyst.* **158,** 744–752 (2002).
24. Lassak, T. M., Fouch, M. J., Hall, C. E. & Kaminski, E. Seismic characterization of mantle flow in subduction systems: can we resolve a hydrated mantle wedge? *Earth Planet. Sci. Lett.* **243,** 632–649 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Anderson for providing slab contours of the southern Andean slab and insights on possible deformation processes in the Andean mantle wedge. We thank S. Pozgay and D. Wiens for discussions on the origin of trench-parallel anisotropy in the Mariana subduction system. We also thank G. Abers and E. Syracuse for discussions on three-dimensional slab geometry and kinematics. This research was supported by the National Science Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to E.A.K. (ekneller@umich.edu).

METHODS

Model set-up. Our modelling approach closely follows the two-dimensional kinematic-dynamic models of ref. 25 but modified for three dimensions[20]. The model domain is a cartesian box (Supplementary Fig. 1) and is divided into four parts: (1) a kinematic slab, (2) a 40-km-thick rigid overriding plate, (3) a rigid wedge corner, and (4) a viscous mantle wedge (Supplementary Fig. 1). At the side boundaries with constant x-coordinate we impose symmetry boundary conditions, which imply zero normal velocity, zero heat flow and zero tangential stress. The effect of symmetry boundary conditions on finite strain calculations was tested by systematically moving their locations. The symmetry boundary conditions used in this work are adjusted so that they do not significantly affect finite strain calculations. The inflow and outflow boundaries are modelled as stress-free boundaries, much as they were in ref. 25. The geometries used for the segments of the Mariana and Andean subduction zones are shown in Supplementary Fig. 1b and c.

Thermomechanical model. Fluid flow in the mantle wedge is governed by the conservation of momentum and mass for an incompressible infinite Prandtl number fluid and is driven by velocity boundary conditions that are parallel to the surface of the subducting slab. The magnitude of convergence velocity is set equal to 5 cm yr^{-1}. We ignore local buoyancy effects in the wedge. We have found that this approximation is appropriate for the employed dry rheological law based on a comparison of two-dimensional models with buoyancy forces, because the local Rayleigh number for secondary convection is near critical. Although we have observed transient effects, the steady-state flow geometry associated with these two-dimensional models is very similar to cases where flow is driven by the slab. Recent studies show that the addition of large quantities of water (hydrogen point defects) in localized regions[15] and crustal foundering[26] may give rise to small-scale convection in the mantle wedge. However, it is difficult to determine the effects of small-scale convection on fabric development from these studies because integrated finite strain calculations were not performed.

A temperature- and stress-dependent rheology is implemented in the viscous wedge. Specifically, we use the constitutive equation:

$$\varepsilon_{t,ij} = A\exp(-E/RT)\sigma^{(n-1)}\sigma_{ij}$$

where $\varepsilon_{t,ij}$ are the components of the strain-rate tensor, σ_{ij} are components of the stress tensor, σ is the second invariant of the stress tensor, T is the absolute temperature, and R is the gas constant. We use $A = 10^{-11.9}$ s^{-1} Pa^{-3}, $E = 510$ kJ mol^{-1} and $n = 3$, consistent with experiments on natural olivine aggregates[21,22]. We impose a viscosity cut-off of 10^{26} Pa s to simulate low-temperature creep mechanisms that limit the magnitude of deviatoric stress. Temperature is obtained by solving the time-dependent heat advection–diffusion equation with constant thermal conductivity of 3 W mK^{-1}, heat capacity $c_p = 1{,}250$ J kg^{-1} K^{-1} and density $\rho = 3{,}300$ kg m^{-3} for the entire domain. The initial thermal condition is defined with a half-space cooling model with a cooling age of 50 Myr. The model evolution is followed for 20 Myr of subduction. This duration of subduction is sufficient to reach steady state in the mantle wedge. Within the final velocity field we compute particle motion and finite strain using a fourth-order Runge–Kutta scheme. The initial position of tracers is defined by a uniform grid with 5 km spacing. Tracers are also injected into the base of the mantle wedge every 200,000 years. The computed finite strain is interpolated back to the uniform three-dimensional grid. The two-dimensional maps and cross-sections of finite strain (for example, Fig. 2a, b) are produced by spatial averaging with an averaging radius of 10 km to mitigate effects due to uneven particle coverage.

Numerical methods. The finite element package Sepran[27] is used to discretize equations on a grid composed of linear tetrahedra. Strong local refinement is implemented in the corner of the wedge with a minimum resolution of 2 km along vertical planes with constant x-coordinates. Minimum resolution in the direction perpendicular to this plane is 5 km in the nose of the wedge to 10 km away from the regions with strong boundary layers. The maximum size of elements is 15 km. The implicit Euler scheme is used for time integration with a dimensional time step approximately equal to twice the Courant–Friedrichs–Levy criterion.

25. van Keken, P. E., Kiefer, B. & Peacock, S. M. High-resolution models of subduction zones: implications for mineral dehydration reactions and the transport of water into the deep mantle. *Geochem. Geophys. Geosyst.* **3**, doi:10.1029/2001DC000256 (2002).
26. Behn, M. D., Hirth, G. & Kelemen, P. B. Trench-parallel anisotropy produced by foundering of arc lower crust. *Science* **317**, 108–111 (2007).
27. Cuvelier, C., Segal, A. & Van Steenhoven, A. A. *Finite Element Methods and Navier-Stokes Equations* (Reidel Publishing, Dordrecht, 1986).

LETTERS

Allometric degree distributions facilitate food-web stability

Sonja B. Otto[1], Björn C. Rall[1] & Ulrich Brose[1,2]

In natural ecosystems, species are linked by feeding interactions that determine energy fluxes and create complex food webs. The stability of these food webs[1,2] enables many species to coexist and to form diverse ecosystems. Recent theory finds predator–prey body-mass ratios to be critically important for food-web stability[3–5]. However, the mechanisms responsible for this stability are unclear. Here we use a bioenergetic consumer–resource model[6] to explore how and why only particular predator–prey body-mass ratios promote stability in tri-trophic (three-species) food chains. We find that this 'persistence domain' of ratios is constrained by bottom-up energy availability when predators are much smaller than their prey and by enrichment-driven dynamics when predators are much larger. We also find that 97% of the tri-trophic food chains across five natural food webs[7] exhibit body-mass ratios within the predicted persistence domain. Further analyses of randomly rewired food webs show that body mass and allometric degree distributions in natural food webs mediate this consistency. The allometric degree distributions hold that the diversity of species' predators and prey decreases and increases, respectively, with increasing species' body masses. Our results demonstrate how simple relationships between species' body masses and feeding interactions may promote the stability of complex food webs.

Natural food webs are characterized by energy and biomass flows across various trophic levels. Despite the structural complexity of these large networks[8], simple food-chain motifs usefully represent

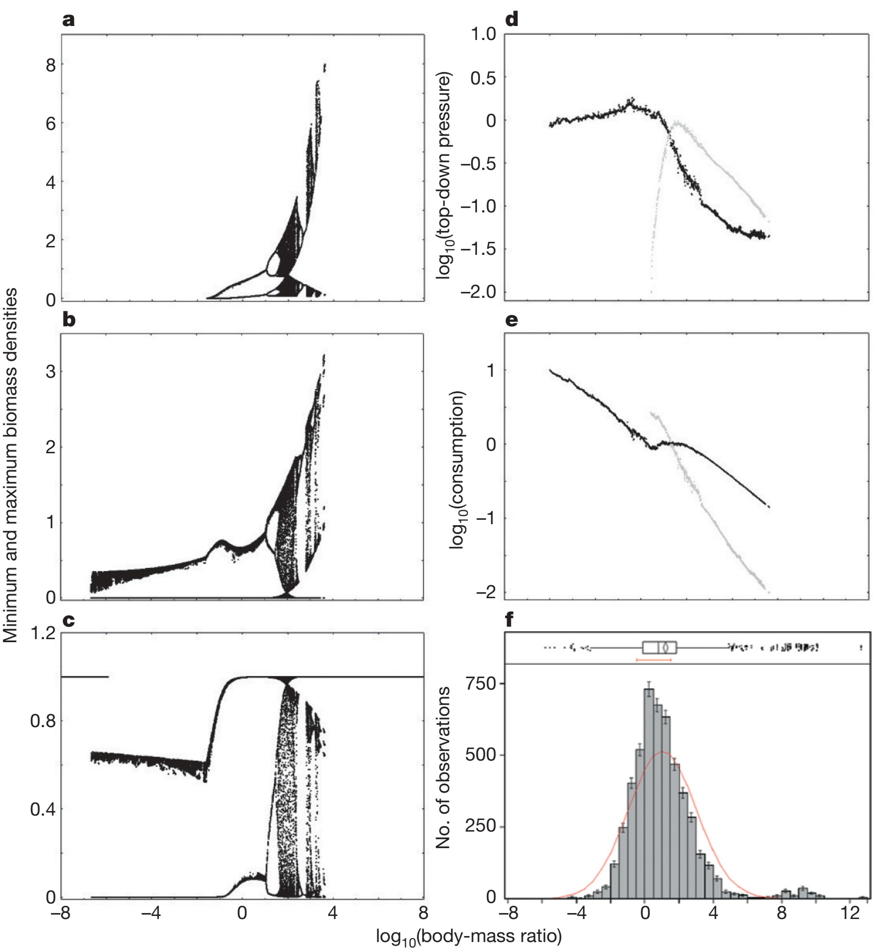


Figure 1 | Population dynamics in tri-trophic food chains. a–c, Effects of R on the biomass minima and maxima of top (**a**), intermediate (**b**) and basal (**c**) species. **d**, Effect of R on $\log_{10}$ of top-down pressure per unit biomass of prey, for intermediate–basal (black) and top–intermediate (grey) species. **e**, Effect of R on $\log_{10}$ of consumption per unit biomass of predator, for intermediate–basal (black) and top–intermediate (grey) species. **f**, Frequency distribution of empirical R in five natural food webs (means ± s.e.m.); the red line shows a normal distribution. An outlier box-plot is shown above the histogram. Simultaneous variation of R of top to intermediate and intermediate to basal species: when $R = 0$, all three species have equal size; when $R < 0$ and $R > 0$, predators are smaller and larger, respectively, than their prey.

[1]Darmstadt University of Technology, Department of Biology, Schnittspahnstrasse 10, 64287 Darmstadt, Germany. [2]Pacific Ecoinformatics and Computational Ecology Lab, Berkeley, California 94703, USA.

the energy transfer[9,10] and mechanisms responsible for non-equilibrium population dynamics in food webs[11–14]. Analyses of food-chain motifs illustrate how population stability under chaotic dynamics may be driven by high resource productivity[11], variation in the species' timescales[14] or certain body-mass ratios between consumers and resources[12]. Population persistence depends on parameters of energy gain (production and consumption) and loss (metabolism and mortality)[15], whose rates per unit biomass follow allometric negative-quarter power-law relationships with the average body masses of the populations[16,17]. We use a bioenergetic model based on these principles[6] to explore how the dynamics of top (t), intermediate (i) and basal (b) species of tri-trophic food chains changes with varying consumer–resource body-mass ratios (R). Our analyses predict the probability of stable coexistence of three invertebrate species in tri-trophic food chains depending on R, which is subsequently evaluated for food chains of five natural food webs[7].

We initially explored a tri-trophic system by simultaneously increasing R between top and intermediate species (R_{ti}) and between intermediate and basal species (R_{ib}) from 10^{-8} to 10^{8} (that is, the consumer is between 10^{8}-fold smaller and 10^{8}-fold larger than its prey). The simultaneous increase in both R values is a simplification to gain knowledge of the population dynamics. The minima and maxima attained for the biomass densities of the three species across this range of R (Fig. 1a–c) depict four distinct stages of coexistence. At the lowest R ($R \leq 10^{-6.7}$), the system exhibits a stable equilibrium where only the basal species persists. At higher R ($10^{-6.7} \leq R < 10^{-1.6}$), two stable attractors appear: either the basal species persists at equilibrium, or basal and intermediate species exhibit globally attractive limit cycles[14]. In this range of R, the top species is much smaller than its prey, and its mass-specific metabolic rate exceeds the energy available from consuming the intermediate species, which prevents persistence[15]. Increasing R above these low ratios decreases the metabolic rates per unit biomass of top and intermediate species and increases the intermediate species' biomass density until the top species' consumption exceeds its metabolic demand enough for the top species to persist ($R = 10^{-1.6}$). Further increases in R ($10^{-1.6} < R < 10^{3.5}$) increases top-down pressure on the intermediate species and decreases top-down pressure on the basal species (Fig. 1d). Increasing R within this range also increases the consumption rate per unit biomass of the intermediate species over that of the top species (Fig. 1e). This counterintuitive result is explained by the simultaneous decrease in the density of intermediate species and increase in the density of basal species, which enhances the energy availability per unit biomass to the intermediate species. This availability increases with R, leading to accelerating oscillations of top and intermediate species (Fig. 1a–c). Mechanistically similar to the 'paradox of enrichment'[18], the dynamics are driven from equilibrium through a series of bifurcations to more complex dynamics until the minimum density of the intermediate species drops below a critical extinction threshold, eliminating both consumer species ($R = 10^{3.5}$; Fig. 1). The complex dynamics in this range of R are caused by the different timescales of the three populations[14]. Further increases in R ($R > 10^{3.5}$) cause unstable dynamics that continue to prevent the persistence of the intermediate and top species (Fig. 1a–c). The persistence of all three species is thus bounded by energy availability to the top species at low R and by enrichment-driven instability of the intermediate species towards higher R.

With this mechanistic background on food-chain dynamics, we decoupled R of upper and lower trophic levels and independently varied both R_{ti} and R_{ib} between 10^{-6} and 10^{13}. This range corresponds to the range of empirical R values of the five natural food webs studied here (Fig. 1f). In 19.6% of this parameter space, we found persistence of all three species (Fig. 2, red areas). The energy-availability boundary of this persistence domain depends on R_{ib}, which needs to exceed a threshold ($R_{ib} > 10^{-1.6}$) within a broad range of R_{ti} ($R_{ti} > 10^{-4.3}$) to increase the density of the intermediate species (that is, the energy available) enough for the top species to persist (Fig. 2, left boundary of red areas). If R_{ib} and R_{ti} exceed a second threshold, both top and intermediate species cease to persist as a result of enrichment-driven dynamics (Fig. 2, right boundary of red areas). This enrichment boundary is determined more continuously and interactively by both R_{ib} and R_{ti} than the energy-availability boundary (Fig. 2).

The persistence domain in Fig. 2 implies that a tri-trophic food chain with R randomly chosen from the range $10^{-6} \leq R \leq 10^{13}$ has a 19.6% chance of persisting. However, 97.5% ± 4.1% (mean ± s.d.) of all invertebrate tri-trophic food chains across five natural food webs from different ecosystem types (see Methods) fall within the persistence domain (Fig. 2a, black points; Fig. 2d, black bars). This difference in probabilities clearly suggests that species' body-mass

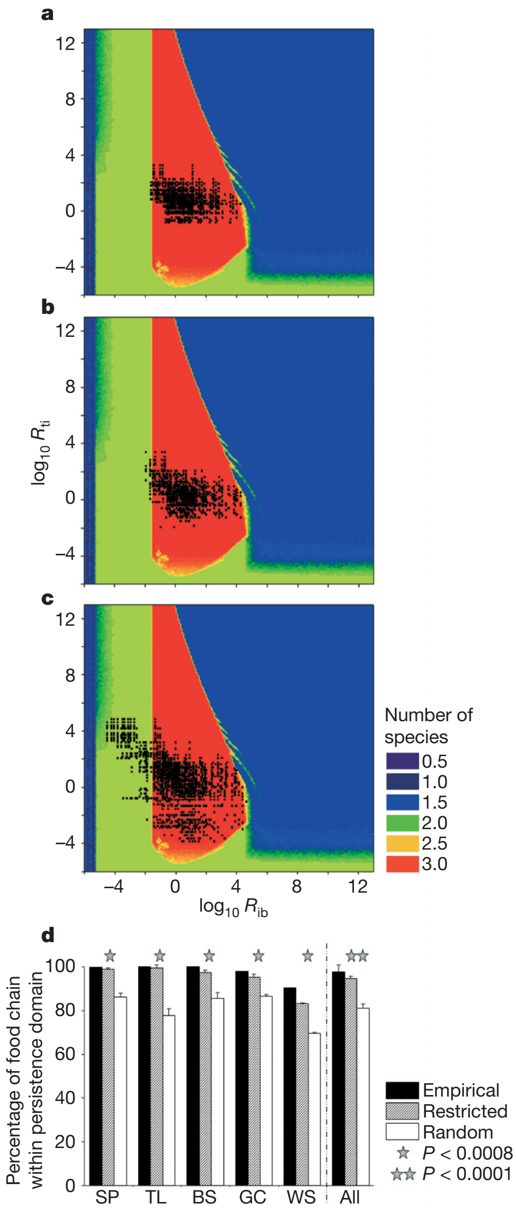


Figure 2 | Population persistence in tri-trophic food chains depending on R_{ti} and R_{ib}. a–c, Colours indicate the numbers of persistent species; red areas characterize a tri-trophic 'persistence domain'. Black points represent food chains of Skipwith pond under empirical food web structures (**a**), restricted rewiring (**b**) and random rewiring (**c**). **d**, Percentages of food chains within the persistence domain (SP, Skipwith pond; TL, Tuesday lake; BS, Broadstone stream; GC, Grand Cariçaie; WS, Weddell Sea) under empirical structures, restricted and random rewiring; results are shown as means and s.d. Stars indicate significant differences between the rewired versions of each food web.

distributions in these food webs strongly stabilize food-chain dynamics. To further explore this hypothesis, we randomly rewired the empirical food webs in a way that preserves the body masses of the species and the total number of links while completely disrupting the food-web topology ('random rewiring'; see Methods). An average of 81.0% ± 7.0% (mean ± s.d.) of these rewired food chains in each of the five food webs fell within the persistence domain (Fig. 2c, d, white bars). This probability is 4.1-fold the 19.6% probability of food chains with randomly distributed body masses within empirically observed ranges that are systematically and independently linked. However, 81% is significantly lower than the 97.5% probability that empirical food chains overlap with the persistence domain ($P < 0.01$). This difference suggests that, while the distribution of species' body masses found in natural food webs provides a substantial increase in the dynamical stability of possible food chains, topological properties of actual food chains might further facilitate food-web stability. To explore which topological properties can provide this additional stabilization, we tested whether correlations between the body mass and degree of species (that is, the number of predator and prey links of a population) drive this effect. To do this, we randomly rewired the food webs with a second randomization algorithm that preserves the body mass and degree of each species ('restricted rewiring'; see Methods). An average of 94.7% ± 6.2% (mean ± s.d.) of the food chains rewired in this restricted way lie within the persistence domain (Fig. 2b, d, grey bars). This probability is 4.8-fold the probability of food chains with randomly distributed body masses (19.6%) and differs significantly from randomly rewired networks (81.0%, $P < 0.05$), but it is not significantly lower than that in empirical food chains (97.5%, $P > 0.17$).

Overall, our results suggest that the distributions of and correlations between the body mass and degree of species within food webs are important mechanisms responsible for food-chain stability. Other topological properties of food webs seem to be of more minor importance. Instead, preserving allometric degree distributions realizes probabilities of tri-trophic stability similar to those found in empirical food webs. This conclusion seems qualitatively insensitive to variation in model parameters (see Supplementary Information). In the five natural food webs studied, the critically important mass–degree relationships are characterized by significant decreases in the number of predator links and significant increases in the number of prey links with increasing body masses of species (Table 1). These simple relationships were removed in the random procedure and retained in the restricted-rewiring procedure (Table 1). Our results seem to reveal a mechanistic basis of body-mass effects on population persistence in simple tri-trophic food chains. Scaling up our analyses to complex food webs suggests that population persistence there could be determined by similar constraints (see Supplementary Information). Although domains of stability using other functional responses also need to be explored, our results for the most widely used nonlinear functional response are of broad importance to ecology. Future extensions of our approach need to also address more variation between network models, species numbers and metabolic types of species to illuminate the generality of the results described here.

Community stability is known to be critically dependent on the body-mass distribution within food webs[3–5]. Here we explore potential mechanisms behind these stability effects by simulating tri-trophic food chains whose persistence is possible under a limited combination of species' body masses that describe a persistence domain. These mechanisms include energy limitation of the top species when predators are much smaller than their prey, and unstable enrichment-driven dynamics of intermediate species when they are much larger. Tri-trophic food chains are frequently parts of more complex motifs within food webs[9,10] that may exhibit more stable dynamics[13,19] or gain additional stability if large top predators couple either spatially separated food chains or other fast and slow energy channels[20–22]. Although ignoring such additional model complexity, the persistence domain predicted by our food-chain model is matched surprisingly well by 97.5% empirical food chains across five natural food webs. Further work on more complex food-web motifs is needed to obtain a better understanding of how body-mass-dependent population persistence scales up with system size from food chains to food webs.

Body masses impose physical constraints on who can hunt, handle and ingest whom in a community[23,24], which determines the diet breadth and foraging behaviour of individual species and topological food-web parameters[4,25,26]. To these relationships between body size and food webs, our study adds allometric degree distributions in which larger species feed on more prey and are consumed by fewer predators than small species. Our study provides a possible explanation for how these distributions may affect characteristics such as population persistence and food-web stability in natural communities. This connection between community-level degree distributions[27,28] and population biology suggests a fundamental bridge between food-web structure[8,28,29] and food-web dynamics[4,5]. Our results illuminate an allometric mechanism that may help to maintain the critically important biodiversity of natural ecosystems.

Table 1 | Allometric degree distributions: dependence of species' link structures (y) on body mass (x)

Food web	Topology	y	Regression equation	R^2	n	P
Skipwith pond	Empirical	No. of predators	$y = -1.00\log x + 5.91$	0.21	33	0.007
		No. of prey	$y = 2.47\log x + 21.01$	0.26	33	0.003
	Random	No. of predators	$y = -0.26\log x + 9.13$	0.05	33	0.20
		No. of prey	$y = -0.08\log x + 9.94$	0.004	33	0.71
Tuesday lake, 1984	Empirical	No. of predators	$y = -0.19\log x + 1.57$	0.47	25	0.0002
		No. of prey	$y = 0.71\log x + 12.61$	0.35	25	0.0017
	Random	No. of predators	$y = -0.03\log x + 3.50$	0.01	25	0.58
		No. of prey	$y = 0.03\log x + 4.30$	0.007	25	0.69
Broadstone stream	Empirical	No. of predators	$y = -0.80\log x - 1.91$	0.40	29	0.0003
		No. of prey	$y = 1.44\log x + 17.76$	0.15	29	0.04
	Random	No. of predators	$y = 0.31\log x + 7.85$	0.10	29	0.10
		No. of prey	$y = -0.24\log x + 3.02$	0.10	29	0.10
Grand Cariçaie, ClControl2	Empirical	No. of predators	$y = -0.54\log x + 4.39$	0.13	102	0.0002
		No. of prey	$y = 0.61\log x + 11.59$	0.05	102	0.03
	Random	No. of predators	$y = -0.06\log x + 7.36$	0.006	102	0.45
		No. of prey	$y = -0.05\log x + 7.44$	0.004	102	0.51
Weddell Sea shelf	Empirical	No. of predators	$y = -0.44\log x + 16.93$	0.02	275	0.03
		No. of prey	$y = 1.96\log x + 20.64$	0.10	275	<.0001
	Random	No. of predators	$y = 0.04\log x + 17.68$	0.002	275	0.50
		No. of prey	$y = 0.01\log x + 17.64$	0.0003	275	0.79

Linear least-square regressions of the number of predators and prey per species (y) on the $\log_{10}$ body masses (x) of the species of five food webs under empirical food-web structures and randomly rewired networks. Empirical networks and restricted rewired networks (not shown) show similar degree distributions, because the restricted rewiring algorithm preserves the number of predators prey per species; n is the number of invertebrate species in the food web.

METHODS SUMMARY

Simulations. A bioenergetic population dynamics model[6] defines the biomass evolution, dB/dt, of basal (b), intermediate (i) and top (t) species:

$$dB_b/dt = r_b G_b B_b - x_i y_i F_{ib} B_i / e \quad (1a)$$

$$dB_i/dt = -x_i B_i + x_i y_i F_{ib} B_i - x_t y_t F_{ti} B_t / e \quad (1b)$$

$$dB_t/dt = -x_t B_t + x_t y_t F_{ti} B_t \quad (1c)$$

where e is the assimilation efficiency, G_b is the logistic net growth (with carrying capacity K) and F is a type II functional response. Biological rates r, x and y (growth, metabolism and maximum consumption) scale with body mass, M: $r,x,y \propto M^{-0.25}$. r, x, y were normalized to the growth rate of basal species (thus, $r_b = 1$), and y was normalized to x. The maximum consumption rate was constant ($y = 8$); x increased with the body-mass ratio to basal species:

$$x_{i,t} = a\left(\frac{M_{i,t}}{M_b}\right)^{-0.25}$$

where a is a constant ($a = 0.2227$ when top, intermediate and basal species are invertebrates[16]). $R_{ti} = M_t/M_i$ and $R_{ib} = M_i/M_b$ are varied between 10^{-8} and 10^{13}, which influences their rates of metabolism (x) and consumption (xyF) per unit biomass. Initial biomass densities were random, simulations were run over 100,000 time steps. Maximum and minimum biomass densities of persistent populations ($B > 10^{-30}$) were recorded, and combinations of persistent R_{ti} and R_{ib} were defined as a 'persistence domain'. The averages of the top-down pressure per unit biomass on basal and intermediate species are $P_{b,i} = x_{i,t} y_{i,t} F_{ib,ti} B_{i,t} / B_{b,i}$, and the energy fluxes per unit biomass to intermediate and top species are $E_{i,t} = x_{i,t} y_{i,t} F_{ib,ti}$.

Rewiring. We compared R_{ti} and R_{ib} of the persistence domain with those of all tri-trophic food chains from five natural food webs (see Table 1)[7]. We created two additional versions of each empirical food web under random and restricted rewiring. For each treatment we calculated the fraction of food chains that were located within the persistence domain of our simulations under three conditions: empirical food web structures, restricted rewiring and random rewiring. Differences in results were evaluated by independent Mann–Whitney U-tests. Relationships between the degree and body mass of species were analysed by ordinary linear least-square regressions.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 13 July; accepted 12 October 2007.

1. De Ruiter, P. C., Wolters, V., Moore, J. C. & Winemiller, K. O. Food web ecology: Playing Jenga and beyond. *Science* **309**, 68–70 (2005).
2. Montoya, J. M., Pimm, S. L. & Solé, R. V. Ecological networks and their fragility. *Nature* **442**, 259–264 (2006).
3. Emmerson, M. C. & Raffaelli, D. Predator–prey body size, interaction strength and the stability of a real food web. *J. Anim. Ecol.* **73**, 399–409 (2004).
4. Loeuille, N. & Loreau, M. Evolutionary emergence of size-structured food webs. *Proc. Natl Acad. Sci. USA* **102**, 5761–5766 (2005).
5. Brose, U., Williams, R. J. & Martinez, N. D. Allometric scaling enhances stability in complex food webs. *Ecol. Lett.* **9**, 1228–1236 (2006).
6. Yodzis, P. & Innes, S. Body size and consumer–resource dynamics. *Am. Nat.* **139**, 1151–1175 (1992).
7. Brose, U. *et al.* Body sizes of consumers and their resources. *Ecology* **86**, 2545 (2005).
8. Williams, R. J. & Martinez, N. D. Simple rules yield complex food webs. *Nature* **404**, 180–183 (2000).
9. Bascompte, J. & Melian, C. J. Simple trophic modules for complex food webs. *Ecology* **86**, 2868–2873 (2005).
10. Milo, R. *et al.* Network motifs: Simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
11. Hastings, A. & Powell, T. Chaos in a three-species food chain. *Ecology* **72**, 896–903 (1991).
12. Jonsson, T. & Ebenman, B. Effects of predator–prey body size ratios on the stability of food chains. *J. Theor. Biol.* **193**, 407–417 (1998).
13. McCann, K., Hastings, A. & Huxel, G. R. Weak trophic interactions and the balance of nature. *Nature* **395**, 794–798 (1998).
14. Muratori, S. & Rinaldi, S. Low- and high frequency oscillations in three-dimensional food chain systems. *SIAM J. Appl. Math.* **52**, 1688–1706 (1992).
15. Gard, T. C. Persistence in food webs: Holling type II food chains. *Math. Biosci.* **49**, 61–67 (1980).
16. Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M. & West, G. B. Toward a metabolic theory of ecology. *Ecology* **85**, 1771–1789 (2004).
17. Savage, V. M., Gillooly, J. F., Brown, J. H., West, G. B. & Charnov, E. L. Effects of body size and temperature on population growth. *Am. Nat.* **163**, E429–E441 (2004).
18. Rosenzweig, M. L. Paradox of enrichment: destabilization of exploitation of ecosystems in ecological time. *Science* **171**, 385–387 (1971).
19. Fussmann, G. F. & Heber, G. Food web complexity and chaotic population dynamics. *Ecol. Lett.* **5**, 394–401 (2002).
20. Koelle, K. & Vandermeer, J. Dispersal-induced desynchronization: from metapopulations to metacommunities. *Ecol. Lett.* **8**, 167–175 (2005).
21. McCann, K. S., Rasmussen, J. B. & Umbanhowar, J. The dynamics of spatially coupled food webs. *Ecol. Lett.* **8**, 513–523 (2005).
22. Rooney, N., McCann, K., Gellner, G. & Moore, J. C. Structural asymmetry and the stability of diverse food webs. *Nature* **442**, 265–269 (2006).
23. Woodward, G. *et al.* Body size in ecological networks. *Trends Ecol. Evol.* **20**, 402–409 (2005).
24. Brose, U. *et al.* Consumer–resource body-size relationships in natural food webs. *Ecology* **87**, 2411–2417 (2006).
25. Beckerman, A. P., Petchey, O. L. & Warren, P. H. Foraging biology predicts food web complexity. *Proc. Natl Acad. Sci. USA* **103**, 13745–13749 (2006).
26. Jonsson, T., Cohen, J. E. & Carpenter, S. R. Food webs, body size, and species abundance in ecological community description. *Adv. Ecol. Res* **36**, 1–84 (2005).
27. Montoya, J. M. & Solé, R. V. Topological properties of food webs: from real data to community assembly models. *Oikos* **102**, 614–622 (2003).
28. Stouffer, D. B., Camacho, J., Guimera, R., Ng, C. A. & Amaral, L. A. N. Quantitative patterns in the structure of model and empirical food webs. *Ecology* **86**, 1301–1311 (2005).
29. Cattin, M. F., Bersier, L. F., Banasek-Richter, C., Baltensperger, R. & Gabriel, J. P. Phylogenetic constraints and adaptation explain food-web structure. *Nature* **427**, 835–839 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank U. Jacob for providing the Weddell Sea data; A. de Roos, E. Berlow, S. Scheu and M. Visser for comments; R. Williams for simulation programs; and N. Martinez for editorial assistance. Financial support was provided by the German Research Foundation.

Author Contributions S.B.O., B.C.R. and U.B. contributed equally to this work. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.B.O. (sonotto@bio.tu-darmstadt.de).

METHODS

Model. Population dynamics of three invertebrate species in a food chain follows a bioenergetic model[6] of the biomass evolution (see equations (1)). $G_b = 1 - B_b/K$ and $F_{ib} = B_b/(B_0 + B_b)$; $F_{ti} = B_i/(B_0 + B_i)$ with a half saturation density B_0. Here, the fraction of the biomass removed from the resource population that is actually eaten is set to unity, which is often characterized as the mechanistically simplest model of predator–prey interactions[30]. The biological rates of production (W), metabolism (X) and maximum consumption (Y) follow negative-quarter power-law relationships with the species' body masses[16]:

$$W_b = a_r M_b^{-0.25} \tag{2a}$$

$$X_{i,t} = a_x M_{i,t}^{-0.25} \tag{2b}$$

$$Y_{i,t} = a_y M_{i,t}^{-0.25} \tag{2c}$$

where a_r, a_x and a_y are allometric constants[6]. The timescale of the system is defined by setting the mass-specific growth rate to unity (equation (3a)). Then the mass-specific metabolic rates of all species, x, are normalized by the timescale (equation (3b)), and the maximum consumption rates, y, are normalized by the metabolic rates:

$$r_i = 1 \tag{3a}$$

$$x_{i,t} = \frac{X_{i,t}}{W_b} = \frac{a_x}{a_r}\left(\frac{M_{i,t}}{M_b}\right)^{-0.25} \tag{3b}$$

$$y_{i,t} = \frac{Y_{i,t}}{X_{i,t}} = \frac{a_y}{a_x} \tag{3c}$$

Substituting equations (3a)–(3c) into equations (1a) and (1b) yields a population dynamic model with allometrically scaled and normalized parameters.

Here the body mass of the basal species, M_b, is set to unity, and the body masses of all other species, M_i and M_t, are expressed relative to the body mass of the basal species. This makes the results presented here independent of the body mass of the basal species.

Simulations. In simulations of tri-trophic food chains, the R values between the top and intermediate species (R_{ti}) and between the intermediate and basal species (R_{ib}) define the body masses M_i and M_t. We used constant values for the other model parameters: maximum ingestion rate $y_{i,t} = 8$ for invertebrate predators; assimilation efficiency $e = 0.85$ for carnivores; carrying capacity $K = 1$; half saturation density of the functional response $B_0 = 0.5$; allometric constant $a = a_x/a_r = 0.2227$ (top, intermediate and bottom species were simulated as invertebrates). We sought a mechanistic explanation for the influence of R on food-web stability by simulating food chains as the simplest multitrophic motif with energy transfer across several trophic levels. This characterizes complex natural food webs better than bitrophic consumer–resource relationships. Analyses of more complex motifs such as omnivory modules require knowledge about the relative interaction strengths of generalist predators with their multiple prey, which was not available for the natural food webs studied.

We varied R between the top and intermediate species ($R_{ti} = M_t/M_i$) and between the intermediate and basal species ($R_{ib} = M_i/M_b$) between 10^{-8} and 10^{13}, which decreased their rates of metabolism (x) and consumption (xyF) per unit biomass. Simulations started with uniformly random biomass densities ($0.05 < B_{t,i,b}\,(T = 0) < 1$) and ran more than 100,000 time steps (T) or until the largest species attained two biomass minima. We recorded the maximum and minimum biomass densities in the second half of the time series of the persistent populations ($B > 10^{-30}$) and defined a 'persistence domain' of combinations of R_{ti} and R_{ib} that enabled persistence of the three populations. For every time series we calculated the averages of the top-down pressure per unit biomass on the basal species, $P_b = x_i y_i F_{ib} B_i/B_b$, and the energy flux per unit biomass to the intermediate species, $E_i = x_i y_i F_{ib}$. Similar calculations yield the averages of the top-down pressure per unit biomass on the intermediate species and the energy flux per unit biomass to the top species.

Evaluation and rewiring. Subsequently, we compared the R_{ti} and R_{ib} values of the persistence domain with those of all tri-trophic food chains across five natural food webs: one from a stream (Broadstone stream), one from a pond (Skipwith pond), one from a lake (Tuesday lake, 1984), one terrestrial (Grand-Cariçaie, ClControl2) and one marine (Weddell Sea shelf) from a global data base[7]. To allow comparisons with our simulations, we studied only food chains of three invertebrate species that composed the vast majority of food chains in the empirical food webs, whereas few food chains include vertebrates or plant species. To test our hypotheses we created two additional versions of each of these empirical food webs under random and restricted rewiring. The 'random rewiring' algorithm conserved only the species' body masses and the total number of links, n, of the empirical food webs and randomly relinked n species pairs without any restrictions. The 'restricted rewiring' algorithm (see ref. 10 and references therein) randomly selects two predator–prey pairs and reconnects the predator of the first pair with the prey of the second pair and vice versa. This rewiring required that none of the new links already existed and ensured the conservation of the total number of predators and prey of each species along with their body masses and the total number of links in the network. We relinked n pairs of links in each food web 20 times to create a random rewired version of the network. Each of the two algorithms was applied to each of the five food webs studied with eight replicates. For each replicate we calculated the fraction of invertebrate food chains with body-mass ratios that were located within the persistence domain of our simulations under three conditions: empirical food web structures, restricted rewiring and random rewiring.

Statistics. Differences in these fractions between the three versions of the food webs were statistically evaluated by eight independent Mann–Whitney U-tests. In each test the five empirical probabilities were tested against five probabilities for each rewiring algorithm (randomly drawn from the eight replicates for each food web). Subsequently, each test was characterized by the highest of the eight P values. The relationships between the numbers of predator links and prey links and the body masses of the species were analysed by ordinary linear least-square regressions. Regressions were performed for each empirical replicate and one randomly rewired replicate of each of the five food webs.

30. Jeschke, J. M., Kopp, M. & Tollrian, R. Predator functional responses: Discriminating between handling and digesting prey. *Ecol. Monogr.* 72, 95–112 (2002).

LETTERS

Nanog safeguards pluripotency and mediates germline development

Ian Chambers[1], Jose Silva[2,3], Douglas Colby[1], Jennifer Nichols[2,4], Bianca Nijmeijer[1], Morag Robertson[1], Jan Vrana[1], Ken Jones[2,4], Lars Grotewold[1] & Austin Smith[2,3]

Nanog is a divergent homeodomain protein found in mammalian pluripotent cells and developing germ cells[1,2]. Deletion of *Nanog* causes early embryonic lethality[2], whereas constitutive expression enables autonomous self-renewal of embryonic stem cells[1]. Nanog is accordingly considered a core element of the pluripotent transcriptional network[3–7]. However, here we report that Nanog fluctuates in mouse embryonic stem cells. Transient downregulation of Nanog appears to predispose cells towards differentiation but does not mark commitment. By genetic deletion we show that, although they are prone to differentiate, embryonic stem cells can self-renew indefinitely in the permanent absence of Nanog. Expanded *Nanog* null cells colonize embryonic germ layers and exhibit multilineage differentiation both in fetal and adult chimaeras. Although they are also recruited to the germ line, primordial germ cells lacking Nanog fail to mature on reaching the genital ridge. This defect is rescued by repair of the mutant allele. Thus Nanog is dispensible for expression of somatic pluripotency but is specifically required for formation of germ cells. Nanog therefore acts primarily in construction of inner cell mass and germ cell states rather than in the housekeeping machinery of pluripotency. We surmise that Nanog stabilizes embryonic stem cells in culture by resisting or reversing alternative gene expression states.

The presence of Nanog is considered a hallmark of pluripotent cells *in vivo* and *in vitro*, and loss of Nanog an early marker of differentiation. In culture, constitutive expression of Nanog can enhance embryonic stem-cell self-renewal and maintain an uncommitted state under conditions in which pluripotency would otherwise be lost[1,8]. However, Nanog is undetectable in a fraction of cells that express Oct4 (Fig. 1a)[9]. To investigate whether mosaic distribution of Nanog is determined at the level of transcription, we used homologous recombination to insert *eGFP* at the *Nanog* AUG codon (Supplementary Figs 1 and 2). Two targeted clones (TNG-A and TNG-B) gave qualitatively similar results, described below. Cultures exhibited a broad range of fluorescence, with up to 20% of cells lacking detectable green fluorescent protein (GFP) (Fig. 1b). Most cells in both GFP$^+$ and GFP$^-$ populations expressed the undifferentiated embryonic stem-cell markers SSEA1 and Oct4. Immunoblotting confirms that GFP expression reflects Nanog protein expression (Fig. 1c). Furthermore, immunofluorescent staining shows that a subset of Oct4$^+$ cells lacks both GFP and detectable Nanog protein (Fig. 1d). This correlation is not absolute, probably because of differences between Nanog and GFP half-lives. After fluorescence-activated cell sorting (FACS) purification, GFP$^+$ cells regenerated cultures containing GFP$^-$ cells. More significantly, GFP$^-$ cells gave rise to GFP$^+$ cells (Fig. 1b). To exclude the possibility that this was due to contamination of the GFP$^-$ population with GFP$^+$ cells, we seeded individual SSEA1$^+$ cells from GFP$^+$ and GFP$^-$ populations (greater than 99% purity) into separate wells of 96-well dishes. GFP$^+$ cells arose from single GFP$^-$ cells at high frequency such that after 7 days most colonies contained GFP$^+$ cells (Supplementary Fig. 3). To visualize re-expression of GFP directly, sorted SSEA1$^+$, GFP$^-$ cells were monitored by time-lapse microscopy. In fields of uniformly GFP$^-$ cells, fluorescent signals appeared by 24 h after plating and continued to emerge thereafter. These cells retained GFP fluorescence during subsequent cell divisions (Fig. 1e and Supplementary Videos).

Not only can GFP$^-$ cells convert to a GFP$^+$ status, but they can expand to form macroscopic colonies containing undifferentiated cells (Supplementary Fig. 3). However, quantification of the colony assays reveals that GFP$^-$ cells form proportionately fewer wholly undifferentiated colonies (Fig. 1f). Overall colony number is also reduced, possibly because of early differentiation with loss of proliferative capacity, although reduced substrate attachment or viability may contribute. Increased differentiation is also seen at high density. Between 1 and 2 days after sorting, GFP$^-$ cells show a significant decrease in the SSEA1$^+$ fraction not observed with GFP$^+$ cells (Supplementary Fig. 4). Collectively these data indicate that embryonic stem-cell cultures contain cells with little or no Nanog expression that are predisposed to differentiate but which may remain undifferentiated and can subsequently re-express Nanog.

We therefore investigated the consequences of definitive elimination of Nanog from embryonic stem cells (Fig. 2). To achieve efficient conditional deletion we initially used ROSA:Cre-ER^{T2} (RC) embryonic stem cells that express tamoxifen-inducible Cre recombinase (Cre-ER^{T2})[10] constitutively from the *ROSA26* locus. A loxP flanked *Nanog* transgene[1] was introduced to create RCN(anog) embryonic stem cells (Supplementary Fig. 1 summarizes the strategies used in this study). Immunoblot analysis indicated that RCN cultures express approximately twice the wild-type Nanog level (Supplementary Fig. 6a). Two rounds of homologous recombination at the *Nanog* locus then generated embryonic stem cells in which the transgene was the sole source of Nanog (Fig. 2 and Supplementary Fig. 5). Treatment with tamoxifen resulted in GFP expression, indicating deletion of *Nanog* sequence (Fig. 2b) with Nanog messenger RNA (mRNA) being eliminated by 48 h (Supplementary Fig. 6b). Previous analysis suggested that removal of Nanog from embryonic stem cells would cause differentiation[2]. Increased differentiation was not initially apparent in tamoxifen-treated cultures (Supplementary Fig. 6c). Only after 5–7 days did overtly differentiated cells appear. However, we found that undifferentiated GFP$^+$ cells persisted upon continuous passaging. If selection was applied for drug resistance expressed from the *Nanog* locus, differentiated cells could be ablated

[1]MRC Centre Development in Stem Cell Biology, Institute for Stem Cell Research, School of Biological Sciences, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh EH9 3JQ, UK. [2]Wellcome Trust Centre for Stem Cell Research, [3]Department of Biochemistry, and [4]Department of Physiology, Development and Neuroscience, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK.

and pure populations of embryonic stem cells readily propagated. However, with time, GFP^{-} cells began to predominate. We surmised that these GFP^{-} cells expanded from initially rare cells that had not deleted *Nanog*. Therefore, FACS-purified GFP^{+} cells were plated at clonal density and stable expandable GFP^{+} cultures were obtained by picking individual colonies. Immunoblotting confirmed that Nanog protein was absent in these cells. In contrast, Oct4 and Sox2 proteins continued to be expressed (Fig. 2c). *Nanog*$^{-/-}$ cells show morphological features of embryonic stem cells, although close inspection reveals a flatter shape with more prominent cytoplasmic protrusions, typical of embryonic stem cells on the periphery of colonies (Fig. 2d and Supplementary Fig. 6). Polymerase chain reaction with prior reverse transcription (RT–PCR) showed maintained expression of all embryonic stem-cell-associated transcripts (ECATs)[2] except Nanog itself (Fig. 2e and data not shown). Even at high cycle number, no expression from *Nanog* or *Nanog* retrogenes[11] was detected (Supplementary Fig. 6d). The presence of Stella and Rex1 mRNAs indicates that Nanog null embryonic stem cells have not converted into epithelial epiblast stem cells (EpiSCs)[12,13]. *Nanog*$^{-/-}$ cells expand more slowly than wild-type cells (data not shown). However, *Nanog*$^{-/-}$ cells have extensive multiplication capacity. During a continuous 3-month culture period, *Nanog*$^{-/-}$ cells retained a modal chromosome count of 40.

For quantitative assessment of self-renewal potential, we used colony-forming assays to compare *Nanog*$^{-/-}$ cells with parental cells (both *Nanog*$^{+/+}$ and *Nanog*$^{+/-}$) from which the *Nanog* transgene had also been excised by treatment with tamoxifen (Fig. 2f and Supplementary Fig. 7). Upon deletion of *Nanog* alleles there is a stepwise reduction in both the total number of colonies formed and the proportion of colonies expressing alkaline phosphatase. A reciprocal increase is observed both in the number of differentiated colonies and the number of colonies containing cells of a primitive endodermal morphology (Supplementary Fig. 7). Crucially, however, all *Nanog*$^{-/-}$ cell lines can form undifferentiated alkaline-phosphatase-positive colonies. The generation of such colonies is wholly dependent upon stimulation with leukaemia inhibitory factor (LIF), confirming that Nanog null embryonic stem cells are not EpiSCs, which are unresponsive to LIF[12,13]. This observation provides further confirmation that self-renewal mechanisms directed by LIF[14] can operate independently of Nanog[15].

To exclude the possibility that exposure to the *Nanog* transgene induced some permanent modification enabling subsequent Nanog independent propagation, we engineered a null genotype in E14Tg2a cells by direct Cre deletion of the endogenous gene. The first *Nanog* allele was inactivated by conventional gene targeting and the second was modified by placing loxP sites in the 5′ and 3′ untranslated regions (UTRs) (Supplementary Fig. 8). Subsequent Cre transfection replaced the *Nanog* coding sequence with GFP, allowing FACS purification of cells that had undergone the deletion. *Nanog*$^{-/-}$ embryonic stem cells were established from two independent, conditionally targeted, embryonic stem cell lines. These cells lacked Nanog transcripts and protein (Supplementary Fig. 8d, e) but could be continuously propagated, retaining the capacity to form undifferentiated colonies at clonal density (Supplementary Fig. 8f).

The differentiation of *Nanog*$^{-/-}$ embryonic stem cells was assayed using the defined neural induction protocol[16,17]. Both RCN-derived and conditionally deleted null cells underwent neural differentiation as measured by loss of Rex1 and Oct4 (Supplementary Fig. 9a) and upregulation of β-III tubulin (Supplementary Fig. 9b). GATA4 and GATA6 were also upregulated compared with parental cells, indicating that the normal block to primitive endoderm differentiation under these conditions is released (Supplementary Fig. 9a). Indeed GATA4 and GATA6 transcripts were detectable in *Nanog*$^{-/-}$ cultures, consistent both with the presence of apparently primitive endodermal cells in the colony-forming assays and with previous reports that absence of Nanog is permissive for hypoblast

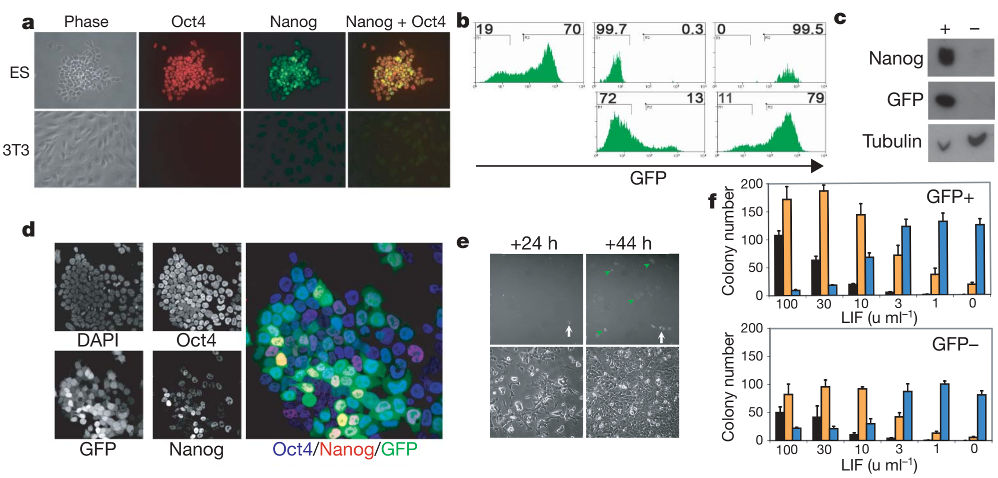


Figure 1 | Nanog expression within the undifferentiated embryonic stem-cell population is reversible. **a**, Nanog is undetectable by immunofluorescence in a subset of undifferentiated embryonic stem cells that express Oct4. Panels show E14Tg2a embryonic stem cells and fibroblast cells that do not express Nanog mRNA (3T3). **b**, Embryonic stem cells with GFP targeted to Nanog (TNG cells) were profiled by flow cytometry (top left). GFP^{+} and GFP^{-} cells were then separated (top centre and right) and after culture for 6 days, FACS analysis was repeated; numbers are the percentage of cells in each of the indicated gates. **c**, Immunoblot of TNG cells for Nanog and GFP. After FACS sorting to greater than 90% purity, cells from SSEA1^{+}, GFP^{+} and SSEA1^{+}, GFP^{-} fractions were lysed and analysed by immunoblotting; +, GFP^{+} fractions; −, GFP^{-} fractions. **d**, Immunofluorescence of TNG cells for Oct4 and Nanog. Individual signals from 4,6-diamidino-2-phenylindole (DAPI), GFP, anti-Oct4 and anti-Nanog are shown on the left alongside a combined view of GFP with the stainings from anti-Oct4 and anti-Nanog. **e**, Individual GFP^{+} TNG cells arise from fields of GFP^{-} TNG cells. After FACS sorting to greater than 99.5% purity for lack of GFP expression, SSEA1^{+} cells were re-plated and imaged at 24–44 h after plating. The start and end images of the bright field and fluorescent series are presented. The white vertical arrow indicates a cell that has activated GFP expression before the start of image collection; this cell divides twice during the next 20 h. Green arrowheads indicate some of the cells that activate synthesis of GFP during the course of image collection. Full movies are in Supplementary Data. **f**, GFP^{-} TNG cells have a reduced capacity to self-renew at clonal density. TNG cells were sorted for SSEA1 expression, separated into GFP^{-} (more than 97% pure) and GFP^{+} (more than 93% pure) populations and assayed as described in Methods. Black bars indicate undifferentiated colonies, blue bars indicate differentiated colonies and orange bars indicate colonies with a mixture of undifferentiated and differentiated cells. Data from a representative experiment are shown; error bars are standard deviations ($n = 3$).

differentiation[2]. When injected into the kidney capsule, *Nanog*$^{-/-}$ cells produced teratomas containing tissues representative of the three germ layers (Supplementary Fig. 10).

These data establish that *Nanog* can be deleted entirely from embryonic stem cells without extinguishing self-renewal or multilineage differentiation potential. *Nanog*$^{-/-}$ embryonic stem cells maintain an undifferentiated morphology, are clonogenic, continue to express undifferentiated markers and are pluripotent. To determine whether they retain the capacity for incorporation into the developing embryo, aggregation chimaeras between *Nanog*$^{-/-}$ cells and wild-type morulae were generated. Advantage was taken of the constitutive GFP expression in RCN-derived *Nanog*$^{-/-}$ embryonic stem cells to track contribution in chimaeras. At E12.5, GFP^{+} cells are distributed throughout the fetus (Fig. 3a). Tissue sections show that *Nanog*$^{-/-}$ cells are integrated into neuroepithelia (Fig. 3b), liver (Fig. 3c) and heart muscle (Fig. 3d). Furthermore, contributions to multiple post-natal chimaeras were evident by coat-colour and GFP expression (Fig. 3e–g and Supplementary Fig. 10g). *Nanog*$^{-/-}$ cells generated by conditional gene targeting also produced coat-colour chimaeras (Supplementary Fig. 9g). Continued Nanog expression is therefore neither required for the capacity of embryonic stem cells to be incorporated into the epiblast nor for the maintenance of somatic lineage potential and mature differentiation.

Nanog is expressed in primordial germ cells throughout the period of epigenetic erasure and germ-cell commitment[18]. To examine whether *Nanog*$^{-/-}$ embryonic stem cells were capable of contributing to the germ line, genital ridges were dissected from chimaeras and stained for expression of Oct4 and Mvh. Oct4 is expressed in germ cells

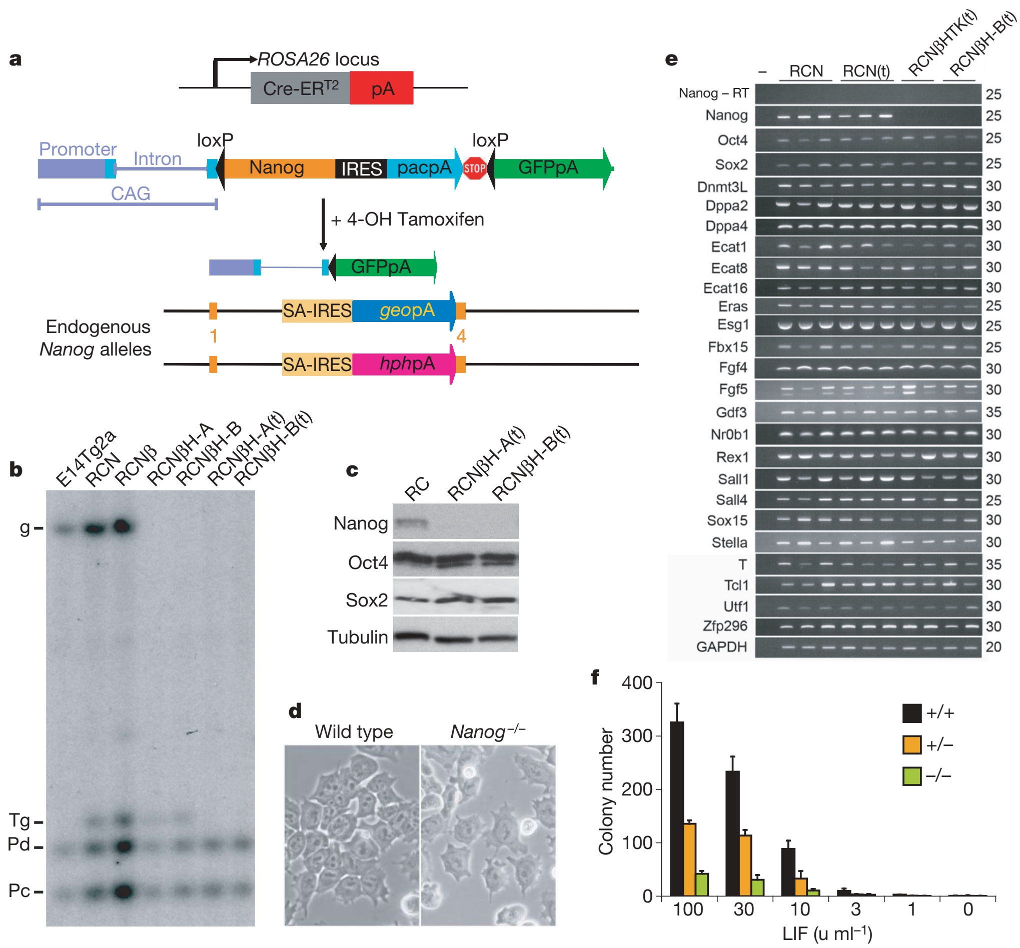


Figure 2 | *Nanog*$^{-/-}$ embryonic stem cells retain expression of pluripotency markers and the capacity for *in vitro* self renewal. a, Genetic manipulation of Nanog expression. E14Tg2a cells were modified by introduction of a Cre-ER^{T2} expression cassette into the *ROSA26* locus to produce RC cells. A Cre-revertible *Nanog* transgene was introduced, producing RCN cells. The endogenous *Nanog* genes were then inactivated by targeted replacement of sequences from within intron I to the 3′ UTR (see Supplementary Fig. 5 for details). A *β-geo*-based vector was used to produce heterozygous RCNβ cells, and *hph*- and *hph-tk*-based vectors were then used to produce nullizygous RCNβH and RCNβTK cells. (The relation of the various cell lines used in this study is summarized in a flow diagram in Supplementary Fig. 1.) After activation of Cre by tamoxifen, the *Nanog* transgene is deleted and the cells become green. The excision, indicated by the suffix '(t)', is illustrated in cells with both Nanog alleles deleted but was also performed in *Nanog*$^{+/+}$ and *Nanog*$^{+/-}$ cells. **b**, Southern analysis using an internal probe derived from the *Nanog* homeodomain. Samples include cells doubly targeted for *Nanog* both before, RCNβH-A, B, and after, RCNβH-A(t), B(t), treatment with tamoxifen. DNA was digested with *Xba*I and *Xho*I, which cut in the Nanog 5′ and 3′ UTRs, respectively. The identities of the bands are indicated: g, the *Nanog* gene produces a band of about 6 kilobases (kb); Tg, the *Nanog* transgene produces a band of about 2.5 kb. The Nanog retrogenes, *NanogPd* and *NanogPc*[11], produce bands of 2.1 and 1.8 kb and provide convenient loading controls. **c**, Immunoblot analysis shows loss of Nanog protein is not accompanied by loss of Oct4 or Sox2 proteins. **d**, Morphology of wild-type and *Nanog*$^{-/-}$ RCNβH-B(t) embryonic stem cells. **e**, RT–PCR of ECATs[2] on RNA from RCN$^{(+/+)}$, RCN$^{(+/+)}$(t), RCNβHTK$^{(-/-)}$(t) and RCNβH-B$^{(-/-)}$(t) embryonic stem cells; the identities of the detected transcripts and the number of PCR cycles used are indicated on the left and right, respectively. Information on primers is in Supplementary Table 1. Each lane shows the analysis of RNA from a separate culture of the indicated cell lines. Control for genomic DNA contamination of RNAs used Nanog primers in complementary DNA (cDNA) reactions lacking reverse transcriptase, because these primers produce similarly sized bands from cDNA and genomic DNA owing to the presence of *Nanog* retrogenes[11]; −, no cDNA. **f**, Clonal assays demonstrate retention of self-renewal capacity. RCN(t), +/+; RCNβ(t), +/−; RCNβH-B(t), −/− cells were assayed as described in Methods Summary. Data are from a representative experiment and show the number of completely undifferentiated colonies formed; error bars are standard deviations ($n = 3$).

by E8.5 (ref. 19) whereas Mvh becomes detectable at E11.5 as primordial germ cells (PGCs) colonize the genital ridge[20]. At E11.5, the expression of Oct4 appeared comparable in GFP$^+$ (*Nanog*$^{-/-}$) and GFP$^-$ (host wild-type) cells. However, Mvh expression was noticeably lower in *Nanog*$^{-/-}$ cells (Fig. 4a, Supplementary Fig. 11 and Supplementary Table 3). By E12.5, no GFP$^+$ cells expressed either Oct4 or Mvh (Fig. 4a, Supplementary Fig. 11 and Supplementary Table 4). This finding indicates that *Nanog*$^{-/-}$ cells persist only in the soma of the genital ridge beyond E11.5 and suggests that Nanog is required for PGCs to prosecute the germ-cell development programme beyond E11.5.

To determine if the defect in PGC development could be assigned unambiguously to the absence of *Nanog*, one of the mutant alleles in the RCNβH-B(t) cells was repaired by homologous recombination (Supplementary Fig. 12). Targeted clones were identified by Southern analysis (Supplementary Fig. 12b). Immunoblotting confirmed that these cells expressed Nanog protein (Supplementary Fig. 12c). Assaying embryonic stem-cell colony-forming efficiency confirmed restored functionality (Supplementary Fig. 12d). In aggregation chimaeras, the corrected cells contributed to the germ lineage at E12.5 (Fig. 4a, Supplementary Fig. 11 and Supplementary Table 4). Therefore the inability of *Nanog*$^{-/-}$ embryonic stem cells to form germ cells is specifically attributable to the lack of *Nanog* expression.

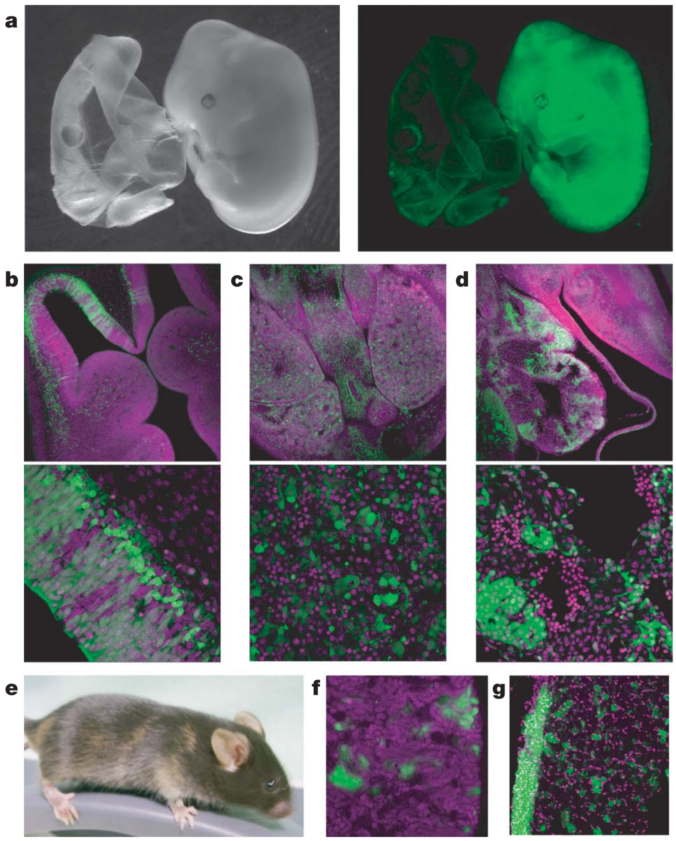


Figure 3 | *Nanog*$^{-/-}$ cells retain the potential for embryo colonization and somatic contribution to chimaeras. **a**, E12.5 fetus from an aggregation of *Nanog*$^{-/-}$, *GFP*$^+$ RCNβH-B(t) cells with GFP$^-$ morula (B6/CBA × B6/CBA: F_2) photographed under normal (left) and fluorescent (right) illumination. **b–d**, E12.5 chimaera sections showing contribution of *Nanog*$^{-/-}$, *GFP*$^+$ RCNβH-B(t) cells to neuroepithelium (**b**), liver (**c**) and cardiac musculature of the left ventricle (**d**). Sections were counterstained with the nuclear dye To-pro-3 (magenta). Original magnifications: top panels ×10; bottom panels ×63. **e**, Post-natal chimaera generated by injection of *Nanog*$^{-/-}$ cells (RCNβH-B(t)) into C57Bl/6 blastocysts. Patches of sandy coat colouring indicate embryonic stem-cell contribution. **f, g**, Sections of adult chimaeras at 3 months showing contribution of *Nanog*$^{-/-}$, *GFP*$^+$ RCNβH-B(t) cells to kidney (**f**) and skeletal muscle (**g**): counterstaining was by autofluorescence at 594 nm (**f**) or by DAPI (**g**).

PGCs colonizing the genital ridge between E11 and E12.5 are undergoing profound epigenetic changes. These include imprint erasure[21], genome-wide demethylation[22] and, in XX cells, reactivation of the inactive X chromosome[23]. Based on the present finding that Nanog is essential for PGCs to complete this transition and the previous demonstration that Nanog is required for the development of the inner cell mass in the blastocyst[2], we propose that the primary function of Nanog is in construction of the unique epigenetically erased states of pluripotent cells and germ cells, respectively.

Earlier studies concentrated on the increased differentiation induced by deletion[2] or short interfering RNA (siRNA) knock down[7] of Nanog and did not test whether embryonic stem cells could in fact persist without Nanog. The existence and properties of Nanog null embryonic stem cells does not contradict these data, but demonstrates that contrary to previous assumptions[15], Nanog does not serve an essential role in conjunction with Oct4 and Sox2 in the transcriptional housekeeping machinery of pluripotency[3,4]. Continuing expression of embryonic stem cell-associated transcript (ECAT) genes in *Nanog*$^{-/-}$ embryonic stem cells calls into question any substantive role for Nanog in co-regulating these and other alleged targets in undifferentiated cells[3,4,24].

Despite expression of all ECAT and other genes hitherto associated with pluripotency, embryonic stem cells lacking Nanog show markedly reduced self-renewal efficiency accompanied by an increased propensity to differentiate into primitive endoderm-like cells.

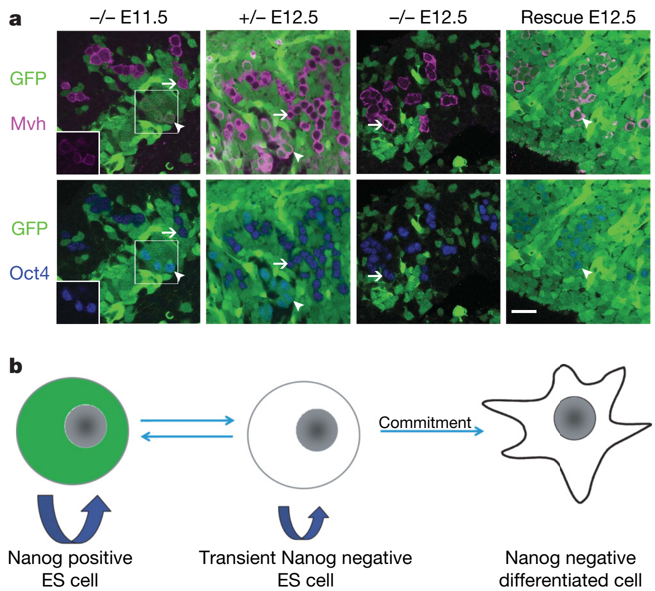


Figure 4 | Nanog is required for cell state transitions during germ cell development and for cell state reversions in embryonic stem-cell cultures. **a**, Genital ridges were dissected from chimaeric mice at the indicated gestational stage and examined by confocal microscopy after double immunofluorescent staining for Mvh and Oct4. Staining for Mvh and Oct4 is shown separately; combined red–green–blue (RGB) staining is shown in Supplementary Fig. 11, as is a tritanope-visible version of the Oct4 panel. Arrows indicate GFP$^-$ (that is, Nanog$^+$), Mvh$^+$, Oct4$^+$ cells and arrowheads indicate GFP$^+$ (that is Nanog$^-$) Mvh$^+$, Oct4$^+$ cells. Chimaeras are from aggregations of RCNβ(t) (+/−), RCNβH−B(t) (−/−) and RCNβH−B(t)R10 (rescue) cells. For clarity, the boxed areas in the E11.5 panels are shown as inserts in the bottom left-hand corner after removal of the GFP signals. Scale bar, 20 μm. **b**, Model of the relation between Nanog expression, self-renewal and differentiation of embryonic stem cells. Embryonic stem cells can express different levels of Nanog and accordingly exhibit a greater or lesser incidence of self-renewal. Commitment to differentiation is a separable event from loss of Nanog expression. However, Nanog acts to safeguard self-renewal by countering the effects of differentiation inducers and preventing progression to commitment, possibly by reversing nascent epigenetic modifications.

However, Nanog null cells do not differentiate exclusively into primitive endoderm, indicating that Nanog does not simply restrict entry into this lineage but rather reduces global differentiation options, as observed when Nanog is overexpressed. We suggest that Nanog regulates the efficiency of self-renewal in a manner analogous to a rheostat (Fig. 4b). In this paradigm, fluctuating Nanog levels in wild-type cultures confer a variable resistance to differentiation upon individual embryonic stem cells. Low levels of Nanog may constitute a 'window of opportunity' in which intrinsic or environmental perturbations can become consolidated into a lineage commitment decision, or may be reversed by re-expression of Nanog. Indeed Nanog has been reported to reverse the expression of a GFP transgene driven by the *brachury* promoter in embryonic stem cells[25]. This may reflect an ability to eliminate pre-commitment changes in gene expression and re-establish a pristine embryonic stem cell state. Deletion of Nanog removes the possibility of reversion. However, self-renewal can continue unless extrinsic conditions favour commitment. Conversely, forced expression of Nanog eliminates the transitional state, thereby shielding the cell from differentiation induction and rendering self-renewal constitutive[1].

It is noteworthy that Nanog is expressed discontinuously in pluripotent cells in the embryo. It is present during formation of pluripotent cells in the blastocyst, is downregulated at the time of implantation[1], then re-expressed in the posterior region of the post-implantation egg cylinder[26]. We speculate that the second wave of Nanog expression may serve to safeguard the egg cylinder epiblast from precocious commitment during gastrulation, in a manner resembling its action in embryonic stem cells. We note that a common feature between the essential functions of Nanog in establishing inner cell mass and germ cells *in vivo* and its ancillary abilities to enhance embryonic stem cell self-renewal and promote nuclear reprogramming[27] is the establishment of cell states with minimal epigenetic governance.

METHODS SUMMARY

Embryonic stem-cell culture. As described[28], *Nanog*$^{-/-}$ cells were maintained by plating at a density of greater than 5×10^4 cells per square centimetre with daily medium changes. For colony-forming assays, embryonic stem cells were trypsinized to a single cell suspension and re-plated at 600 cells per 10 cm^2 well in GMEMβ/10% FCS with the indicated LIF concentration. After incubation for 6 days, plates were stained for alkaline phosphatase and scored for differentiation status.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 September; accepted 22 October 2007.

1. Chambers, I. *et al.* Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643–655 (2003).
2. Mitsui, K. *et al.* The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631–642 (2003).
3. Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
4. Loh, Y. H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genet.* **38**, 431–440 (2006).
5. Wang, J. *et al.* A protein interaction network for pluripotency of embryonic stem cells. *Nature* **444**, 364–368 (2006).
6. Chickarmane, V., Troein, C., Nuber, U. A., Sauro, H. M. & Peterson, C. Transcriptional dynamics of the embryonic stem cell switch. *PLoS Comput. Biol.* **2**, e123 (2006).
7. Ivanova, N. *et al.* Dissecting self-renewal in stem cells with RNA interference. *Nature* **442**, 533–538 (2006).
8. Ying, Q. L., Nichols, J., Chambers, I. & Smith, A. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115**, 281–292 (2003).
9. Hatano, S. Y. *et al.* Pluripotential competence of cells associated with Nanog activity. *Mech. Dev.* **122**, 67–79 (2005).
10. Vallier, L. *et al.* An efficient system for conditional gene expression in embryonic stem cells and in their *in vitro* and *in vivo* differentiated derivatives. *Proc. Natl Acad. Sci. USA* **98**, 2467–2472 (2001).
11. Robertson, M. *et al.* Nanog retrotransposed genes with functionally conserved open reading frames. *Mamm. Genome* **17**, 732–743 (2006).
12. Brons, I. G. *et al.* Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**, 191–195 (2007).
13. Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196–199 (2007).
14. Niwa, H. How is pluripotency determined and maintained? *Development* **134**, 635–646 (2007).
15. Chambers, I. & Smith, A. Self-renewal of teratocarcinoma and embryonic stem cells. *Oncogene* **23**, 7150–7160 (2004).
16. Lowell, S., Benchoua, A., Heavey, B. & Smith, A. G. Notch promotes neural lineage entry by pluripotent embryonic stem cells. *PLoS Biol.* **4**, e121 (2006).
17. Ying, Q. L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nature Biotechnol.* **21**, 183–186 (2003).
18. Yamaguchi, S., Kimura, H., Tada, M., Nakatsuji, N. & Tada, T. Nanog expression in mouse germ cell development. *Gene Expr. Patterns* **5**, 639–646 (2005).
19. Scholer, H. R., Dressler, G. R., Balling, R., Rohdewohld, H. & Gruss, P. Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. *EMBO J.* **9**, 2185–2195 (1990).
20. Toyooka, Y. *et al.* Expression and intracellular localization of mouse Vasa-homologue protein during germ cell development. *Mech. Dev.* **93**, 139–149 (2000).
21. Szabo, P. E., Hubner, K., Scholer, H. & Mann, J. R. Allele-specific expression of imprinted genes in mouse migratory primordial germ cells. *Mech. Dev.* **115**, 157–160 (2002).
22. Hajkova, P. *et al.* Epigenetic reprogramming in mouse primordial germ cells. *Mech. Dev.* **117**, 15–23 (2002).
23. Monk, M. & McLaren, A. X-chromosome activity in foetal germ cells of the mouse. *J. Embryol. Exp. Morphol.* **63**, 75–84 (1981).
24. Shi, W. *et al.* Regulation of the pluripotency marker Rex-1 by Nanog and Sox2. *J. Biol. Chem.* **281**, 23319–23325 (2006).
25. Suzuki, A. *et al.* Maintenance of embryonic stem cell pluripotency by Nanog-mediated reversal of mesoderm specification. *Nature Clin. Pract. Cardiovasc. Med.* **3** (suppl. 1), S114–S122 (2006).
26. Hart, A. H., Hartley, L., Ibrahim, M. & Robb, L. Identification, cloning and expression analysis of the pluripotency promoting Nanog genes in mouse and human. *Dev. Dyn.* **230**, 187–198 (2004).
27. Silva, J., Chambers, I., Pollard, S. & Smith, A. Nanog promotes transfer of pluripotency after cell fusion. *Nature* **441**, 997–1001 (2006).
28. Smith, A. G. Culture and differentiation of embryonic stem cells. *J. Tissue Cult. Methods* **13**, 89–94 (1991).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to V. Karwacki, A. Waterhouse, R. Wilkie, R. MacLay and J. Ure for technical assistance, to C. Manson, J. Verth and colleagues for animal husbandry, and to V. Wilson for comments on the manuscript. This research was supported by the Wellcome Trust, the Juvenile Diabetes Research Foundation, the Medical Research Council and the Biotechnological and Biological Sciences Research Council of the United Kingdom, and a Human Frontier Science Program Fellowship (to L.G.).

Author Contributions I.C. designed the experimental strategy and analysed the data; J.S., J.N. and A.S. contributed to the experimental design. J.S., J.N. and K.J. performed the chimaera study, and J.S. the confocal analyses. D.C. conducted gene targeting and cell biological analysis, and together with J.V. ran the FACS experiments. M.R. and B.N. performed molecular biological analyses. L.G. produced and characterized the ROSA26-Cre-ERT2 cells. I.C. and A.S. conceived the study and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to I.C. (ichambers@ed.ac.uk).

METHODS

DNA constructions. Standard DNA recombination methods were used. The structures of the DNA substrates for homologous recombination are outlined; full sequence information is available on request. For introduction of GFP to the *Nanog* locus homology arms extended 3,640 base pairs (bp) upstream and 3,408 bp downstream from the initiator ATG codon. Between them was placed eGFP-loxP-frt-IRES-pac-frt-SPA-MAZ[29]-loxP. For simple knock-out vectors, the 5′ homology arm extended from the *Xba*I site 72 bp upstream to the *Msc*I site 3,408 bp downstream of the initiator ATG codon, and the 3′ arm from the *Bam*HI site 308 bp upstream to the *Nco*I site 4,289 bp downstream of the polyadenylation site. Between them was placed an IRES-drug resistance-polyadenylation signal cassette for *β-geo*, *hph* or *hph-tk* to produce pgNanog-*geo*, pgNanog-*hph* and pgNanog-*hph-tk*, respectively. For the conditional targeting construct, the 5′ homology arm extended from 3,640 bp upstream to 6,554 bp downstream of the initiation codon, and the 3′ arm extended from the *Pme*I site 29 bp upstream to the *Nco*I site 4,289 bp downstream of the polyadenylation site. In between was placed IRES-pac-C2-MAZ-loxP-eGFP-β-globin pA. A second loxP site was placed in the same orientation at the *Xba*I site in the 5′ UTR.

Embryonic stem-cell culture and cell-line derivation. RC cells were established by homologous recombination such that Cre-ER^{T2} (ref. 10) was expressed from the *ROSA26* locus and the selection cassette was removed by site-specific recombination. RC cells were then transfected with CAGloxPNanog-IRES-pacpA-loxP-eGFPpA[1] to produce RCN cells (see Fig. 2 and Supplementary Fig. 1 for summaries). Targeting *Nanog* with pgNanog-*geo* produced RCNβ cells. The second *Nanog* allele in RCNβ cells was targeted with pgNanog-*hph* or pgNanog-*hph-tk* to give RCNβH and RCNβHTK cells, respectively. Deletion of the *Nanog* transgene, denoted by the suffix '(t)', was achieved by culturing cells in 1 μM 4-OH-tamoxifen. After Cre-mediated deletion, GFP was brought under CAG control, allowing FACS sorting to remove cells that had failed to undergo deletion, and lines were established from the sorted cells.

Lines derived without prior overexpression of Nanog were established by targeting Nanog in E14Tg2a cells with pgNanog-*geo* to give Tβ cells. The second allele in these cells was targeted by using pgNanog-*Conditional* to give TβC lines; these were assessed by PCR of genomic DNA using primers GGGTCACCTTACAGCTTCTTTTGCATTA and GTGATGGCGAGGGAAGGGATTTCTG to identify lines containing the 5′ UTR loxP site. All putative targeted cells were analysed by Southern blotting using probes labelled to greater than 10^9 c.p.m. per microgram according to the strategies described in the relevant Supplementary Figures. Chimaeras were produced by microinjection into C57Bl/6 blastocysts[30] or by morula aggregation[31].

Fluorescence-activated cell analysis. Cells were analysed using a Becton-Dickinson (Palo Alto, California) FACSCalibur flow cytometer. Cells were prepared for FACS sorting by incubation in cell dissociation buffer (Gibco catalogue no 13151-014) for 5 min and resuspension at about 10^6 per millilitre in PBS/10% FCS. Anti-SSEA1 antibody was added to cells at 1:1,000 dilution of ascites, cells incubated at 4 °C for 10 min and washed in PBS/10% FCS. Phycoerythrin-conjugated anti-IgM μ chain antibody was then added at 1:1000 for 10 min at 4 °C and cells sorted on a Dako (Glostrup, Denmark) Cytomation MoFlo High Performance Cell Sorter.

Protein analysis. Nanog was detected by using an affinity-purified rabbit anti-Nanog antibody directed against an amino-terminal epitope (SVGLPGPHSLPSSEE)[32]. Oct4 was detected by using the mouse C10 monoclonal (Santa Cruz). Mvh was detected using Abcam antibody Ab13840. The SSEA1 antibody (MC-480) developed by D. Solter and B. Knowles was obtained from the Developmental Studies Hybridoma Bank, Department of Biological Sciences, University of Iowa, Iowa City, USA. TuJ1 (β-III tubulin) was detected by using a monoclonal antibody (Covance catalogue no. MMS-435P). Immunofluorescence analysis was performed essentially as described[33]. TNG cells were prepared for immunofluorescence analysis by culturing on gelatinized glass slides (superfrost plus) for 48 h. For immunoblot analysis, 10^6 cells were lysed in 100–200 μl of SDS lysis buffer[34]. Lysate (10 μl) was then subjected to SDS–polyacrylamide gel electrophoresis (SDS–PAGE) and transferred to membrane. Blots were then blocked in 10% non-fat dry milk/10 mM Tris pH 7.6/150 mM NaCl/0.05% NP-40 overnight, incubated for 2 h at room temperature with the primary antibody diluted in 5% non-fat dry milk/10 mM Tris pH 7.6/150 mM NaCl/0.15% NP-40 and washed three times in 10 mM Tris pH 7.6/150 mM NaCl/0.5 M NaCl/0.3% Triton X-100. Incubation with secondary antibody was for 1 h in 5% non-fat dry milk/10 mM Tris pH 7.6/150 mM NaCl/0.15% NP-40 followed by three washes in 10 mM Tris pH 7.6/150 mM NaCl/0.5 M NaCl/0.3% Triton X-100. Membranes were then incubated with Super-Signal West Pico (Pierce) for 5 min at room temperature and exposed to Hyperfilm (Amersham).

RNA analysis. Total RNA was prepared using the RNeasy mini kit (Qiagen catalogue no. 74106) with DNase (Qiagen catalogue no. 79254) treatment. Northern analysis was performed by electrophoresis of RNA on 0.66 M formaldehyde, 1% agarose gels and membranes hybridized with probes labelled to approximately 10^9 c.p.m. per microgram against the Nanog open reading frame (ORF) or a fragment of GAPDH. cDNA was primed by $d(T)_n$ on 1 μg RNA using a Superscript kit (Invitrogen catalogue no. 12371-019). Primers for RT–PCR are listed in Supplementary Table 1. RT–PCR was performed on cDNA equivalent to 10 ng total RNA in a 50 μl reaction using Taq polymerase (Qiagen). Ten microlitres of these reactions was used for gel electrophoresis. Q-PCR was performed by using the primers listed in Supplementary Table 2 on a LightCycler 480 (Roche) with cDNA equivalent to 25 ng total RNA. The monocolour hydrolysis probe protocol consisted of denaturation at 95 °C for 5 min followed by 45 cycles of 95 °C, 10 s; 61°C, 10 s with a single data acquisition during each extension cycle.

29. Ashfield, R. *et al.* MAZ-dependent termination between closely spaced human complement genes. *EMBO J.* **13**, 5656–5667 (1994).
30. Schwartzberg, P. L., Goff, S. P. & Robertson, E. J. Germ-line transmission of a c-abl mutation produced by targeted gene disruption in ES cells. *Science* **246**, 799–803 (1989).
31. Nagy, A., Gertsenstein, M., Vintersten, K. & Behringer, R. *Manipulating the Mouse Embryo: A Laboratory Manual* (Cold Spring Harbor Press, New York, 2003).
32. Chambers, I. Mechanisms and factors in embryonic stem cell self-renewal. *Rend. Fis. Acc. Lincei* s.9, v.16 83–97 (2004).
33. Lowell, S., Benchoua, A., Heavey, B. & Smith, A. G. Notch promotes neural lineage entry by pluripotent embryonic stem cells. *PLoS Biol.* **4**, e121 (2006).
34. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685 (1970).

Vol 450 | 20/27 December 2007 | doi:10.1038/nature06385

LETTERS

Isolation of rare circulating tumour cells in cancer patients by microchip technology

Sunitha Nagrath[1]*, Lecia V. Sequist[2]*, Shyamala Maheswaran[2], Daphne W. Bell[2]†, Daniel Irimia[1], Lindsey Ulkus[2], Matthew R. Smith[2], Eunice L. Kwak[2], Subba Digumarthy[2], Alona Muzikansky[2], Paula Ryan[2], Ulysses J. Balis[1]†, Ronald G. Tompkins[1], Daniel A. Haber[2] & Mehmet Toner[1]

Viable tumour-derived epithelial cells (circulating tumour cells or CTCs) have been identified in peripheral blood from cancer patients and are probably the origin of intractable metastatic disease[1–4]. Although extremely rare, CTCs represent a potential alternative to invasive biopsies as a source of tumour tissue for the detection, characterization and monitoring of non-haematologic cancers[5–8]. The ability to identify, isolate, propagate and molecularly characterize CTC subpopulations could further the discovery of cancer stem cell biomarkers and expand the understanding of the biology of metastasis. Current strategies for isolating CTCs are limited to complex analytic approaches that generate very low yield and purity[9]. Here we describe the development of a unique microfluidic platform (the 'CTC-chip') capable of efficient and selective separation of viable CTCs from peripheral whole blood samples, mediated by the interaction of target CTCs with antibody (EpCAM)-coated microposts under precisely controlled laminar flow conditions, and without requisite pre-labelling or processing of samples. The CTC-chip successfully identified CTCs in the peripheral blood of patients with metastatic lung, prostate, pancreatic, breast and colon cancer in 115 of 116 (99%) samples, with a range of 5–1,281 CTCs per ml and approximately 50% purity. In addition, CTCs were isolated in 7/7 patients with early-stage prostate cancer. Given the high sensitivity and specificity of the CTC-chip, we tested its potential utility in monitoring response to anti-cancer therapy. In a small cohort of patients with metastatic cancer undergoing systemic treatment, temporal changes in CTC numbers correlated reasonably well with the clinical course of disease as measured by standard radiographic methods. Thus, the CTC-chip provides a new and effective tool for accurate identification and measurement of CTCs in patients with cancer. It has broad implications in advancing both cancer biology research and clinical cancer management, including the detection, diagnosis and monitoring of cancer[10].

CTCs are rare, comprising as few as one cell per 10^9 haematologic cells in the blood of patients with metastatic cancer, hence their isolation presents a tremendous technical challenge[7,9,11–13]. Microfluidic lab-on-a-chip devices provide unique opportunities for cell sorting and rare-cell detection; they have been successfully used for microfluidic flow cytometry[14], continuous size-based separation[15,16] and chromatographic separation[17]. Despite their success in manipulating microlitre amounts of simple liquids in microscale channels[14,18,19], they have thus far shown limited capability to deal with the cellular and fluid complexity of large volumes (millilitres) of whole blood samples[20–22].

Here we describe the development and application of a microfluidic device (the 'CTC-chip') that can efficiently and reproducibly isolate CTCs from the blood of patients with common epithelial tumours (Fig. 1, and Supplementary Fig. 1). The CTC-chip (Fig. 1b) consists of an array of microposts (Supplementary Fig. 1c) that are made chemically functional with anti-epithelial-cell-adhesion-molecule (EpCAM, also known as TACSTD1) antibodies. Anti-EpCAM provides the specificity for CTC capture from unfractionated blood because EpCAM is frequently overexpressed by carcinomas of lung, colorectal, breast, prostate, head and neck, and hepatic origin, and is absent from haematologic cells[23,24].

Two essential parameters that determine the efficiency of cell capture on the CTC-chip are: (1) flow velocity, because it influences the duration of cell–micropost contact; and (2) shear force, which must be sufficiently low to ensure maximum cell–micropost attachment. To optimize these parameters we employed theoretical analyses characterizing the interaction of cells with microposts distributed within

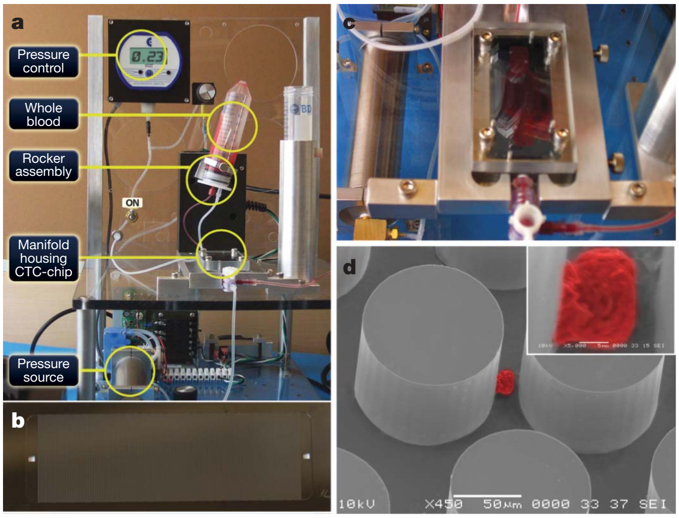


Figure 1 | Isolation of CTCs from whole blood using a microfluidic device. **a**, The workstation setup for CTC separation. The sample is continually mixed on a rocker, and pumped through the chip using a pneumatic-pressure-regulated pump. **b**, The CTC-chip with microposts etched in silicon. **c**, Whole blood flowing through the microfluidic device. **d**, Scanning electron microscope image of a captured NCI-H1650 lung cancer cell spiked into blood (pseudo coloured red). The inset shows a high magnification view of the cell.

[1]Surgical Services and BioMEMS Resource Center, Massachusetts General Hospital, Harvard Medical School, and Shriners Hospital for Children, Boston, Massachusetts 02114, USA. [2]Massachusetts General Hospital Cancer Center, Harvard Medical School, Boston, Massachusetts 02114, USA. †Present addresses: National Human Genome Research Institute/NIH Cancer Genetics Branch, Bethesda, Maryland 20892, USA (D.W.B.); Department of Pathology, University of Michigan Health System, Ann Arbor, Michigan 48109, USA (U.J.B.).
*These authors contributed equally to this paper.

the flow path (Supplementary Figs 1c and 2). Briefly, the simulation indicated an equilateral triangular arrangement, with a 50 μm distance between microposts and a 50 μm shift after every 3 rows, to be the most efficient geometric arrangement. For a given volumetric flow rate of 1 ml h^{-1} through the device, the maximum shear stress experienced by a cell near the micropost surface was estimated to be 0.4 dyn cm^{-2} at $\theta = 68°$ and the expected maximum velocity was 460 μm s^{-1} (Supplementary Fig. 2b–d), within the range facilitating maximum cell attachment according to linear shear stress chamber studies (Supplementary Fig. 3). On the basis of the simulation results, we fabricated a CTC-chip containing an array of 78,000 microposts within a 970 mm^2 surface.

To determine the efficiency of capture, we spiked non-small-cell lung cancer (NSCLC) cells (NCI-H1650) into phosphate buffered saline (PBS) at 100 cells ml^{-1} and captured the spiked cancer cells using the CTC-chip. NSCLC cells were visually evident about EpCAM-coated microposts, whereas no cancer cells were seen following flow through uncoated posts (Supplementary Fig. 4a–c). The calculated capture efficiency was 65% and decreased significantly at flow rates above 2.5 ml h^{-1} (Fig. 2a), presumably owing to increased shear stress, consistent with our simulation predictions. The efficiency of capture was not enhanced at flow rates less than 0.75 ml h^{-1}, leading us to select a flow rate of 1–2 ml h^{-1} for subsequent studies.

To determine the effect of cellular EpCAM expression on efficiency of CTC capture, we compared capture rates among cancer cell lines with varied EpCAM expression, including NSCLC NCI-H1650 cells and breast cancer SkBr-3 cells, with >500,000 antigens per cell; prostate cancer PC3-9 cells, with approximately 50,000 antigens per cell; and bladder cancer T-24 cells, with approximately 2,000 antigens per cell[25]. Each cell line was spiked into PBS at a concentration of 100 cells ml^{-1} and analysed, resulting in a mean capture yield of > 65% in all cases (Fig. 2b). Interestingly T-24 cells with relatively low EpCAM expression were captured as efficiently as high-level antigen-expressing cells. We believe this is due to the augmented cell–substrate interactions inherent within the CTC-chip.

To evaluate cell capture under more physiological conditions, we conducted a series of experiments in which NCI-H1650 cells were spiked into whole blood from healthy donors. Suspensions at concentrations ranging from 50 to 50,000 tumour cells per ml of whole blood were analysed, yielding recovery rates of >60% in all cases (Fig. 2c). To assess the potential steric hindrance of red blood cells in the flow path, these studies were repeated using lysed blood from healthy donors. Capture rates were comparable under both conditions ($r^2 = 0.99$; Fig. 2c). Similar results were obtained using whole blood and lysed samples from NSCLC patients (Supplementary Fig. 4d, e). We thus concluded that the CTC-chip does not require blood-sample pre-processing.

Having optimized the CTC-chip with controlled quantities of cancer-derived cells, we tested its capacity to capture CTCs from whole blood samples donated by cancer patients. A total of 116 samples from 68 patients with epithelial cancers including NSCLC ($n = 55$), prostate ($n = 26$), pancreatic ($n = 15$), breast ($n = 10$) and colon ($n = 10$) were studied. The majority of patients had metastatic disease; however, 7 of 26 subjects with prostate cancer had untreated, clinically localized disease with specimens collected before prostatectomy with curative intent (Supplementary Table 2). The average volume of blood analysed was 2.7 ml per sample (range, 0.9 to 5.1 ml). We also examined samples from 20 healthy individuals (3.0 ± 0.4 ml (mean ± s.d.) of blood per subject) as controls (Supplementary Table 1).

CTCs captured from a group of patient samples were identified using a comprehensive image analysis algorithm, consisting of staining with 4,6-diamidino-2-phenylindole (DAPI) for DNA content, and using rhodamine-conjugated anti-cytokeratin (also known as KERSMCR) antibodies for epithelial cells, and fluorescein-conjugated anti-CD45 antibodies for haematologic cells (Fig. 2d–k, and Supplementary Fig. 5). Cells captured by anti-EpCAM-coated microposts and staining for cytokeratin were scored as CTCs, whereas CD45-positive cells were scored as contaminating normal haematologic cells. The morphologic characteristics exhibited by the captured CTCs were consistent with malignant cells, including large cellular size with high nuclear:cytoplasmic ratios and visible nucleoli (Fig. 2d–g). The mean viability of captured cells was 98.5 ± 2.3% (mean ± s.d.), determined by assessing cell membrane integrity in 10 high-power fields of view per CTC-chip in 4 samples obtained from lung ($n = 2$) and prostate ($n = 2$) cancer patients.

CTCs were identified in 115 of 116 (>99%) patient samples analysed, with the single negative specimen being a small volume sample (0.9 ml) from a patient with colorectal cancer. The number of CTCs isolated ranged from 5 to 1,281 per ml for NSCLC (155 ± 236 (mean ± s.d.) CTCs per ml), 16 to 292 (86 ± 78) for metastatic prostate, 25 to 174 (94 ± 63) among localized prostate cancers, 9 to 831 (196 ± 228) for pancreatic, 5 to 176 (79 ± 52) for breast, and 42 to 375 (121 ± 127) for colorectal cancers (Fig. 3a, b). The identification of CTCs in subjects with clinically localized prostate cancer at numbers approximating those in metastatic prostate cancer patients is a novel finding enabled by the high sensitivity of our technique and warrants further study. The average purity of capture, as defined by the ratio of cytokeratin$^+$ to CD45$^+$ cells, was 52% in NSCLC, 49% in metastatic prostate, 53% in localized prostate, 53% in pancreatic, 60% in breast, and 67% in colon cancers (Fig. 3c). None of the 20 healthy subjects had any identifiable CTCs (Supplementary Table 1). On the basis of these results we calculated the sensitivity (99.1%) and

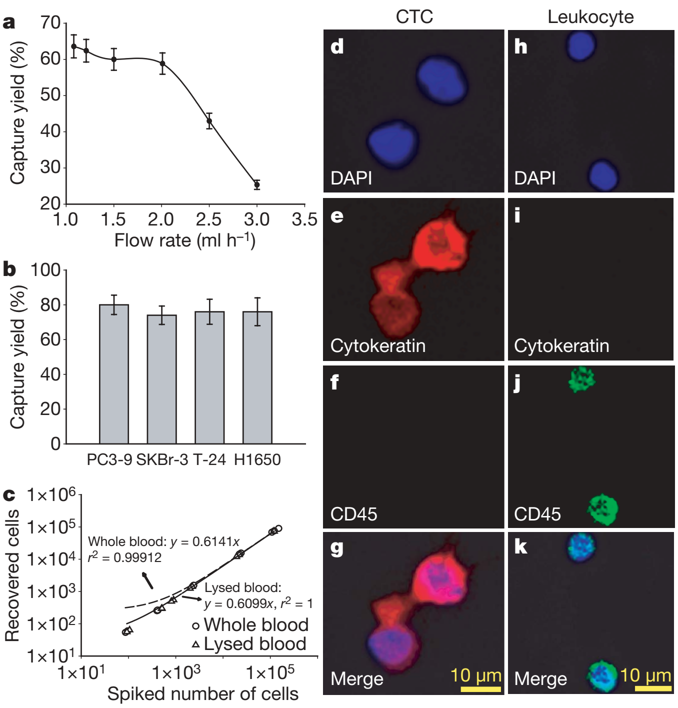


Figure 2 | CTC capture and enumeration. **a**, Capture yield as a function of flow rate. Data shown represent measurements averaged over three devices, and each error bar represents the standard error of the mean. **b**, Capture yields from buffer spiked with 100 cells per ml of four different cell lines: prostate (PC3-9), breast (SkBr-3), bladder (T-24), and NSCLC (NCI-H1650). Each data point was repeated in at least 3 devices. The error bars represent standard deviations of measurements within each experiment. **c**, Regression analysis of capture efficiency for various target cell concentrations, comparing whole blood to lysed blood samples. The plot represents number of cells spiked versus number of cells recovered. **d–k**, Higher magnification (20×) images of captured CTCs and haematologic cells from NSCLC patients, stained with DAPI, and for cytokeratin and CD45. Merged images identify CTCs in panels **d–g** and haematologic cells in panels **h–k**.

specificity (100%) of the CTC-chip across all five cancers. Finally, we evaluated the reproducibility of CTC capture using split samples and showed high experimental reproducibility ($r^2 = 0.98$; Supplementary Fig. 4f).

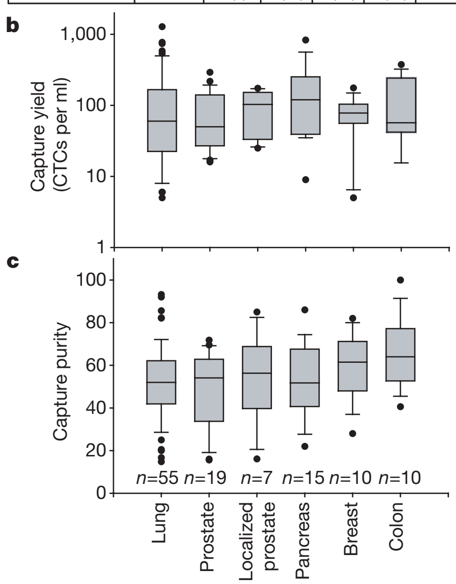
To determine whether captured cells are suitable for subsequent molecular analyses, we tested expression of two tumour-specific markers: prostate-specific antigen (PSA; also known as KLK3) in prostate cancer and thyroid transcription factor-1 (TTF-1; also known as NKX2-1) in lung adenocarcinoma. Specific expression of these markers was evident by immunostaining (Fig. 4a, b, d, e), and confirmed by direct lysis of captured CTCs on the microchip and PCR with reverse transcription (RT–PCR) amplification of the individual transcripts (Fig. 4c, f). Considering the ~50% purity of captured viable CTCs (two orders of magnitude higher than currently available technologies)[26], the CTC-chip provides a powerful opportunity for CTC-based molecular analyses.

To demonstrate the unique clinical potential of our approach, we evaluated the ability of changes in CTC burden to predict changes in tumour volume in cancer patients undergoing anti-cancer treatments. Patients with advanced epithelial-based malignancies that were commencing a first or second line systemic treatment regimen were eligible. Blood samples were collected at baseline and during subsequent clinic visits; the exact follow-up schedule varied between patients. Computed tomograms (CT scans) were performed at baseline and at regular intervals according to standard clinical practice. For each CT scan, we calculated the sum of the unidimensional size in centimetres of all significant and measurable tumour sites, as per the RECIST standardized system[27]. All patients with baseline and at least one follow-up CTC sample and CT scan were analysed, including patients with NSCLC ($n = 3$), colorectal ($n = 2$), pancreatic ($n = 3$) and oesophageal cancer ($n = 1$) (Fig. 3d–i). The absolute number of CTCs captured did not necessarily correspond with tumour size, and may be influenced by other factors. The correlation between per cent change in CTC quantity and per cent change in tumour size, from first to last measurement, was analysed over all 9 patients and yielded a Pearson's correlation coefficient of 0.68 ($P = 0.03$). These results from a small cohort of patients show that CTC quantity correlates reasonably well with clinical response and clinical non-response to treatment (Fig. 3d–i, and Supplementary Fig. 6), indicating that monitoring treatment response using the CTC-chip may ultimately be a powerful tool enabling accurate, early decision-making. The clinical impact of such an approach could be large, and could enable a decrease in patient exposure to toxicities from ineffective therapies and guide them towards the treatments most active against their specific tumour.

Other approaches to enrich or sort CTCs from peripheral blood have been previously published, including flow cytometry[28], immunomagnetic beads[9], high-throughput optical-imaging systems[29], and fibre optic array scanning[12]. Immunomagnetic-bead purification[1,13,30] is currently the lead technology in the clinical setting, with reported success in identifying CTCs in a portion of tested patients with lung, prostate, colon, breast and pancreatic cancer[30]. However, this approach isolates small numbers of CTCs (4 ± 24 (mean ± s.d.) per ml in lung; 11 ± 118 in breast; 10 ± 33 in prostate; and 1 ± 2 in both colorectal and pancreatic cancers)[30] with very low purity (0.01–0.1%)[26], and low yield (~20–60% of patients)[30]. The

a

Cohort	Total no. of samples	Range of CTCs per mL: 5–20	20–50	50–100	>100	Samples with >5 CTC per ml (%)
		No. of samples (%)				
Healthy subjects	20	0	0 (0)	0 (0)	0 (0)	0
Lung cancer	55	11 (20)	11 (20)	11 (20)	22 (40)	100
Prostate cancer	19	3 (16)	8 (42)	1 (5)	7 (37)	100
Localized prostate cancer	7	0 (0)	3 (43)	0 (0)	4 (57)	100
Pancreatic cancer	15	1 (7)	3 (20)	2 (13)	9 (60)	100
Breast cancer	10	2 (20)	0 (0)	4 (40)	4 (40)	100
Colon cancer	10	0 (0)	2 (20)	4 (40)	3 (30)	90



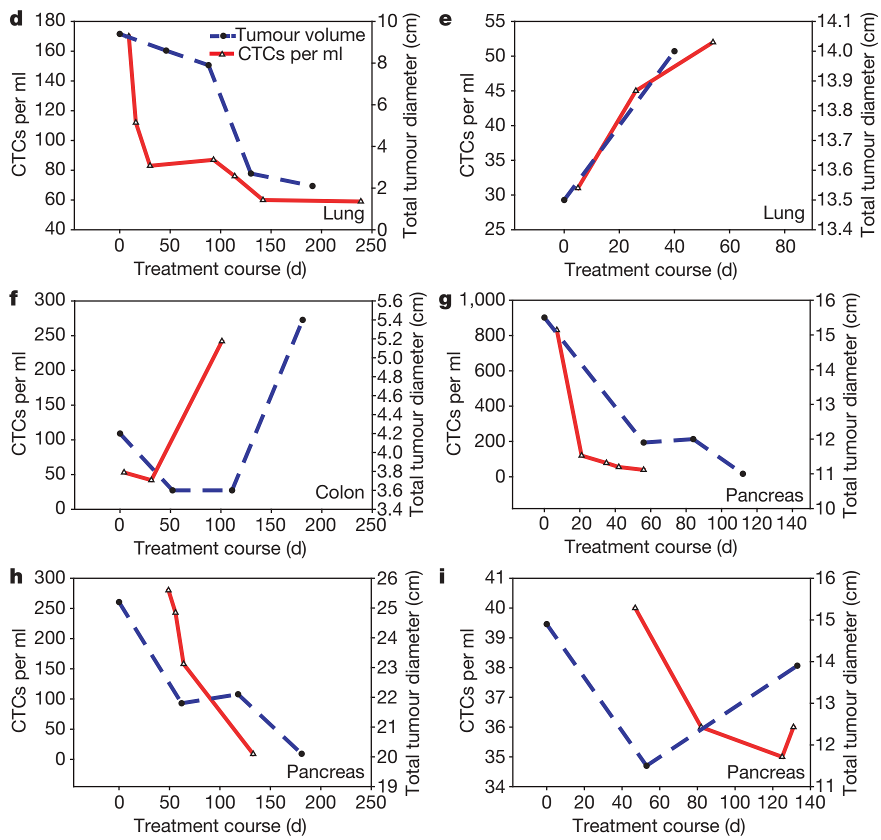


Figure 3 | Enumeration of CTCs from cancer patients. a, Summary of samples and CTC counts per 1 ml of blood in patients with various advanced cancers and localized prostate cancer. **b**, Frequency of CTCs per 1 ml of blood, by diagnosis. The box plot presents the median, lower and upper quartiles (25th, 75th percentiles). Data points that lie outside the 10th and 90th percentiles are shown as outliers. **c**, Purity of captured CTCs (ratio of CTCs to total nucleated cells), by diagnosis. **d–i**, Serial CTC assessment. CTC quantity (red), and tumour size (blue) measured as the unidimensional sum of all significant tumour sites on a CT scan, are well correlated over the course of anti-cancer treatment for nine patients. Six of them are shown here, for whom diagnoses and specific therapies were: NSCLC, 1st-line carboplatin, paclitaxel (**d**); NSCLC, 2nd-line pemetrexed (**e**); colorectal, 1st-line 5FU, irinotecan (**f**); pancreatic, 1st-line gemcitabine, bevacizumab (**g**); pancreatic, 1st-line gemcitabine (**h**); pancreatic, 1st-line gemcitabine, erlotinib (**i**). Baseline CT scans were before therapy initiation and CTC measurements began at or shortly after the first treatment.

level of 'biological noise' associated with the low sensitivity, selectivity, and yield of magnetic-bead-based technologies is prohibitive to their capacity to monitor response to treatment in a dynamic fashion and for early cancer detection. Hence, this method has thus far demonstrated clinical utility only as a gross prognostic tool, classifying patients into high- and low-risk categories[1]. Conversely, the high sensitivity (1 target cell in 1 billion blood cells), selectivity (>47% purity), and yield (99%) of the CTC-chip makes it ideally suited for real-time monitoring of response to cancer therapy.

In addition, the CTC-chip is unique in that it sorts rare cells directly from whole blood in a single step. From a technical perspective, this is possible because the CTC-chip is the first microfluidic device that can successfully process millilitre volumes of whole blood. This contrasts with magnetic-bead-based systems[30] that require multiple 'bulk' semi-automated preparatory steps (centrifugation, washing and incubation), resulting in loss and/or destruction of a significant proportion of the rare cells. In addition to its simplicity, the CTC-chip is readily adaptable for potential use in various clinical scenarios, including changes in throughput and in the antibody on the microposts, allowing capture of other types of rare circulating cells. The CTC-chip's one-step nature and versatility make it conducive to point-of-care use and rapid integration into clinical practice.

Finally, the CTC-chip is distinctive in that its gentle nature (maximum shear stress is 0.4 dyn cm^{-2}) allows for isolation of viable cells, whereas magnetic-bead-based approaches can only isolate fixed, non-viable cells[30]. The stationary nature of the captured cells on fixed microposts allows wash-out of non-specifically bound leukocytes, resulting in a 10^6-fold enrichment, a level of purity that is two orders of magnitude higher than existing technologies. The capacity to isolate concentrated, viable CTCs makes the CTC-chip an ideal tool for molecular access to rare CTC subpopulations such as metastatic precursor cells or cancer stem cells.

In summary, the CTC-chip captures large numbers of viable CTCs in a single step from whole blood without pre-dilution, pre-labelling, pre-fixation or any other processing steps. The techniques described here and the broader application of microfluidic rare-cell capture technology to cancer patients hold significant promise for identifying key biological determinants of blood-borne metastases, and for providing a robust platform aimed at early diagnosis and longitudinal monitoring of cancer.

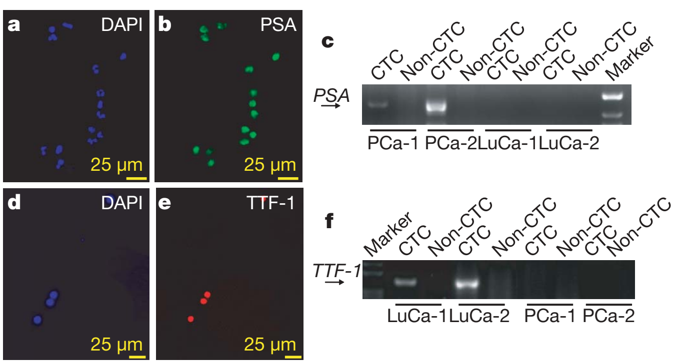


Figure 4 | Characterization of CTCs with tumour-specific molecular markers. **a**, **b**, CTCs from a prostate cancer patient stained positive for DAPI and PSA expression. **c**, RT–PCR amplification of *PSA* transcript is seen in two patients with prostate cancer (PCa), but not in two patients with lung cancer (LuCa), and only in blood fractions enriched for CTCs as opposed to non-enriched fractions (non-CTC). **d**, **e**, CTCs from an NSCLC patient stained for DAPI and TTF-1. **f**, RT–PCR shows expression of *TTF-1* in two patients with lung cancer (LuCa), which is absent in two patients with prostate cancer (PCa), and only when RNA was eluted from blood fractions enriched for CTCs as opposed to non-enriched fractions (non-CTC).

METHODS SUMMARY

The microfluidic system (Fig. 1a) consists of a microfluidic chip etched in silicon (Fig. 1b), a manifold to enclose the chip (Fig. 1c, and Supplementary Fig. 1b), and a pneumatic pump (Fig. 1a) to establish the flow through the capture module (Fig. 1c). The schematic of the microfluidic system is depicted in Supplementary Fig. 1b. The dimensions of the chip are 25 mm × 66 mm, with an active capture area of 19 mm × 51 mm. It contains an equilateral triangular array of microposts, 100 μm tall and 100 μm in diameter with an average 50 μm gap between microposts (Supplementary Fig. 1c). For increased hydrodynamic efficiency, the repeated patterns of equilateral triangular arrays were shifted vertically by 50 μm for every three rows throughout the chip to maximize the interactions between micropost structures and cells. This array incorporates 78,000 microposts within a surface area of 970 mm^2. Microposts were fabricated with deep reactive ion etching (DRIE) by Silex. The blood specimen collection and processing, macro-to-micro coupling, identification and enumeration of CTCs, cell viability and molecular assays are described in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 January; accepted 17 October 2007.

1. Cristofanilli, M. *et al.* Circulating tumor cells, disease progression, and survival in metastatic breast cancer. *N. Engl. J. Med.* **351**, 781–791 (2004).
2. Steeg, P. S. Tumor metastasis: mechanistic insights and clinical challenges. *Nature Med.* **12**, 895–904 (2006).
3. Gupta, G. P. & Massague, J. Cancer metastasis: building a framework. *Cell* **127**, 679–695 (2006).
4. Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105–111 (2001).
5. Mocellin, S., Hoon, D., Ambrosi, A., Nitti, D. & Rossi, C. R. The prognostic value of circulating tumor cells in patients with melanoma: a systematic review and meta-analysis. *Clin. Cancer Res.* **12**, 4605–4613 (2006).
6. Smerage, J. B. & Hayes, D. F. The measurement and therapeutic implications of circulating tumour cells in breast cancer. *Br. J. Cancer* **94**, 8–12 (2006).
7. Rolle, A. *et al.* Increase in number of circulating disseminated epithelial cells after surgery for non-small cell lung cancer monitored by MAINTRAC(R) is a predictor for relapse: A preliminary report. *World J. Surg. Oncol.* **3**, 18 (2005).
8. Braun, S. & Marth, C. Circulating tumor cells in metastatic breast cancer—toward individualized treatment? *N. Engl. J. Med.* **351**, 824–826 (2004).
9. Zieglschmid, V., Hollmann, C. & Bocher, O. Detection of disseminated tumor cells in peripheral blood. *Crit. Rev. Clin. Lab. Sci.* **42**, 155–196 (2005).
10. Bell, D. W. & Haber, D. A. A blood-based test for epidermal growth factor receptor mutations in lung cancer. *Clin. Cancer Res.* **12**, 3875–3877 (2006).
11. Kahn, H. J. *et al.* Enumeration of circulating tumor cells in the blood of breast cancer patients after filtration enrichment: correlation with disease stage. *Breast Cancer Res. Treat.* **86**, 237–247 (2004).
12. Krivacic, R. T. *et al.* A rare-cell detector for cancer. *Proc. Natl Acad. Sci. USA* **101**, 10501–10504 (2004).
13. Racila, E. *et al.* Detection and characterization of carcinoma cells in the blood. *Proc. Natl Acad. Sci. USA* **95**, 4589–4594 (1998).
14. Fu, A. Y., Spence, C., Scherer, A., Arnold, F. H. & Quake, S. R. A microfabricated fluorescence-activated cell sorter. *Nature Biotechnol.* **17**, 1109–1111 (1999).
15. Davis, J. A. *et al.* Deterministic hydrodynamics: taking blood apart. *Proc. Natl Acad. Sci. USA* **103**, 14779–14784 (2006).
16. Huang, L. R., Cox, E. C., Austin, R. H. & Sturm, J. C. Continuous particle separation through deterministic lateral displacement. *Science* **304**, 987–990 (2004).
17. Chang, W. C., Lee, L. P. & Liepmann, D. Biomimetic technique for adhesion-based collection and separation of cells in a microfluidic channel. *Lab Chip* **5**, 64–73 (2005).
18. Whitesides, G. M. The origins and the future of microfluidics. *Nature* **442**, 368–373 (2006).
19. Hong, J. W. & Quake, S. R. Integrated nanoliter systems. *Nature Biotechnol.* **21**, 1179–1183 (2003).
20. Toner, M. & Irimia, D. Blood-on-a-chip. *Annu. Rev. Biomed. Eng.* **7**, 77–103 (2005).
21. Dittrich, P. S. & Manz, A. Lab-on-a-chip: microfluidics in drug discovery. *Nature Rev. Drug Discov.* **5**, 210–218 (2006).
22. El-Ali, J., Sorger, P. K. & Jensen, K. F. Cells on chips. *Nature* **442**, 403–411 (2006).
23. Went, P. T. *et al.* Frequent EpCam protein expression in human carcinomas. *Hum. Pathol.* **35**, 122–128 (2004).
24. Balzar, M., Winter, M. J., de Boer, C. J. & Litvinov, S. V. The biology of the 17–1A antigen (Ep-CAM). *J. Mol. Med.* **77**, 699–712 (1999).
25. Rao, C. G. *et al.* Expression of epithelial cell adhesion molecule in carcinoma cells present in blood and primary and metastatic tumors. *Int. J. Oncol.* **27**, 49–57 (2005).
26. Smirnov, D. A. *et al.* Global gene expression profiling of circulating tumor cells. *Cancer Res.* **65**, 4993–4997 (2005).
27. Therasse, P. *et al.* New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National

Cancer Institute of the United States, National Cancer Institute of Canada. *J. Natl. Cancer Inst.* **92,** 205–216 (2000).
28. Terstappen, L. W. *et al.* Flow cytometry–principles and feasibility in transfusion medicine. Enumeration of epithelial derived tumor cells in peripheral blood. *Vox Sang.* **74** (suppl. 2), 269–274 (1998).
29. Kraeft, S. K. *et al.* Reliable and sensitive identification of occult tumor cells using the improved rare event imaging system. *Clin. Cancer Res.* **10,** 3020–3028 (2004).
30. Allard, W. J. *et al.* Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. *Clin. Cancer Res.* **10,** 6897–6904 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Amin for technical assistance in running experiments, O. Hurtado for clean room work, S. Murthy for surface chemistry, L. Romonosky for cell counting, D. Hyde for digital pictures and D. Poulsen for illustrations. We are also grateful to R. Kapur and his team for technical assistance. The authors acknowledge funding from the National Institutes of Health (to M.T.), and the Doris Duke Distinguished Clinical Scientist Award (to D.A.H.).

Author Contributions S.N., L.V.S., R.G.T., D.A.H. and M.T. designed and conducted the study. S.M., D.W.B. and L.U. performed gene expression analyses; D.I. contributed to the microfluidic system. M.R.S., E.L.K. and P.R. acquired clinical samples. U.J.B. provided input on cytopathology; A.M. performed statistical analysis; and S.D. performed radiology measurements. S.N., L.V.S., D.W.B., S.M., D.I., D.A.H. and M.T. participated in data analysis and writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.T. (mtoner@hms.harvard.edu).

METHODS

Blood specimen collection and processing. Blood samples were drawn from patients with advanced-stage solid tumours before, during and after chemotherapy at Massachusetts General Hospital under an IRB-approved protocol. Blood specimens were also drawn from healthy donors after obtaining informed consent. All specimens were collected into vacutainer tubes containing the anticoagulant EDTA and were processed within 24 h. Between sample collection and sample processing, whole blood specimens were stored at 4 °C on a rocking platform to prevent cell settling. For experiments using lysed blood, NH_4Cl was added to whole blood in 10:1 v/v ratio and mixed for 15–20 min at room temperature. Following centrifugation at 200*g* (10 °C) for 5 min, the supernatant was removed and the pellet re-suspended in an equivalent volume of buffer and stored on a laboratory mixer at 4 °C.

Macro–micro coupling. To establish the flow through the chip, the following methods were adopted to couple the microchip with a macrofluidic pumping system. The silicon chips were purged with nitrogen and sealed with pressure-sensitive adhesive tape (3M, St Paul). The sealed microfluidic devices were then placed in a transparent 5.1 cm × 7.6 cm plastic manifold consisting of a base, top cover and a spacer (Supplementary Fig. 1b). The base has inlet and outlet ports for fluid handling. The manifold base also has six guiding metal pillars, each 1 mm in height, to hold the device in place and in alignment with the inlet and outlet ports. A metal spacer placed between the base and the top cover prevents mechanical stress on the device. The base and top cover attach by screws, providing a leak-proof assembly with minimum dead-volume. For ease of operation the port dimensions are such that standard lure fittings can be used.

A pneumatic macrofluidic drive system was specifically designed to control flow through the microfluidic CTC-chip, as shown in Fig. 1a. It uses a pneumatic pump, pressure regulators, and a digital pressure display to exquisitely control the pressure of the air used to drive blood from a sealed sample container into the CTC-chip. A rocker assembly provides sample mixing throughout the experiment. Before running samples through the chip, the device was purged with 3.0 ml of buffer. A 5 ml aliquot of blood sample was measured into a conical tube, sealed, placed on the rocker unit and connected to the capture device with low dead volume fittings (Fig. 1a). The sample was allowed to mix on the rocker for at least 5 min before running the experiment. The pneumatic pump was turned on and the pressure adjusted according to the required flow rate. After the experiment, the CTC-chip was flushed with 10.00 ml PBS at 10 ml h^{-1} to remove any non-specifically bound cells.

Identification and enumeration of CTCs by fluorescence microscopy. An Olympus SZX (Olympus America) upright reflective microscope with an automated ProScan stage (Prior Scientific) was used to image the microfluidic chip. The microchip was scanned automatically in a 1 mm × 1 mm grid format using the programmable stage and Qcapture Pro software (Media Cybernetics). Captured images (at 100× total magnification) were carefully examined and the objects that met predetermined criteria were counted. Colour, brightness and morphometric characteristics such as cell size, shape and nuclear size were considered in identifying potential CTCs and excluding cell debris and non-specific cells. Cells that stained positive for cytokeratin (cytokeratin$^+$) and met the phenotypic morphological characteristics were scored as CTCs. Cell counts were expressed as the number of cells per 1 ml of blood (CTCs per ml). Cell counts by cytokeratin$^+$ were also confirmed in a group of patients ($n = 8$) using triple stain (cytokeratin$^+$, DAPI$^+$, CD45$^-$). To evaluate inter-operator variability, the same set of scanned images from each chip ($n = 12$) was analysed by two different operators blinded to each other's results. The regression analysis of CTC counts by multiple operators demonstrated correlation coefficient (r^2) of 0.992 with a slope of 1.04, indicating that the percentage of error in counting CTCs from the images is low and that enumeration of cell counts is highly reproducible between operators.

Cell viability. Cell viability was determined with the LIVE/DEAD viability assay kit. This assay is based on intracellular esterase activity of live cells and plasma membrane integrity of dead cells. Briefly, captured CTCs were incubated at room temperature for 30 min in a solution of 2 μM calcein AM and 4 μM ethidium bromide prepared in PBS. At the end of the incubation period, the chip was washed with 1 ml of 1× PBS and visualized under the microscope.

Molecular analysis. Total RNA was isolated from cells using the Picopure RNA isolation kit according to the manufacturer's instruction and subjected to linear amplification using the TransPlex Whole Transcriptome Amplification Kit (Rubicon Genomics). The 509 base pair human PSA coding region was amplified from circulating prostate tumour cell complementary DNA using the following primer pairs (sense and antisense, 5′ to 3′): primary PCR, TTGTGGGAGGCTGGGAGTG and CCTTCTGAGGGTGAACTTGCG; secondary PCR, (GGCAGGTGCTTGTGGCCTCTCGTGG and GTCATTGGAAATAACATGGAGGTCC). The *TTF-1* transcript was amplified using the following primer pairs (sense and antisense, 5′ to 3′): primary PCR, CTGCAACGGCAACCTGGGCAACATG and CAGGTACTTCTGTTGCTTGAAGCG; secondary PCR, CAGGACACCATGAGGAACAGCGCCTC and CAGGTACTTCTGTTGCTTGAAGCG.

Modelling and theoretical device optimization. Micropost geometry and the arrangement of the micropost array were systematically explored in the process of designing the CTC-chip. Three different micropost distributions and arrangements were tested: a square array, a diagonal square array and an equilateral triangular array. The area occupied by the microposts for the square and triangular distribution is given by:

For a square array,

$$\varepsilon = \frac{\pi a^2}{l^2} \tag{1}$$

For a triangular array,

$$\varepsilon = \frac{2\pi a^2}{\sqrt{3} l^2} \tag{2}$$

Where a is the radius of the micropost and l is the centre-to-centre distance between adjacent microposts. Hence, for a given spacing between microposts, an increase in the radius of the micropost increases the micropost density, resulting in higher capture area.

The hydrodynamic efficiency for each distribution is calculated on the basis of the analytical solution derived by Drummond and Tahir[31] and is determined as the ratio of the effective capture length to the mean spacing between the microposts. The effective capture length is based on the limiting trajectory for classical interception by a cylinder calculated from the stream function. Accordingly, for an equilateral triangular distribution of microposts, the flow distribution is given by[31]:

$$U = \frac{F}{8\pi\mu}\left\{ ln\left(\frac{1}{\varepsilon}\right) - 1.497504972 + 2\varepsilon - \frac{\varepsilon^2}{2} - 0.739137296\varepsilon^4 + \frac{2.534185018\varepsilon^5}{1 + 1.275793652\varepsilon} \right\} \tag{3}$$

Where U is the mean velocity between microposts, F is the force per unit length acting on the cell, μ is the fluid viscosity, and ε is as given by equation (2).

We evaluated the efficiency of cell capture using the three different micropost array distributions and found that the calculated hydrodynamic efficiency of capture as a function of the spacing between the microposts was greatest with the equilateral triangular micropost arrangement, and improved with shorter centre-to-centre spacing between the microposts (Supplementary Fig. 1a). However, excessive reduction in micropost spacing might lead to physical trapping of cells between the microposts; therefore, we selected a 50 μm distance between microposts as the calculated optimal spacing, which yields a theoretical capture efficiency of 65%. Consequently, a detailed computational analysis of the hydrodynamics (flow distribution and cell trajectories) for an equilateral triangular array distribution of microposts was performed using commercially available finite element microfluidics solver COMSOL.

To ensure the CTC-chip would be sensitive and allow for high-throughput, we employed numerical analyses using COMSOL to characterize the interaction of cells with the microposts distributed within the flow path. Two factors essential to achieve high CTC capture are: (1) optimization of flow velocity to maximize frequency of contact between cells and the chemically modified microposts, and (2) optimization of shear forces to ensure that they are lower than those favouring cell attachment to the posts. Micropost size, spacing and the distribution along the streamlines are the critical variables that determine flow velocity and shear stress. Hence, we modelled the flow properties and cell–micropost interactions for an equilateral triangular array (Supplementary Fig. 1b). We found that the triangular micropost arrangement resulted in non-linear streamlines, facilitating cell contact with micropost surfaces. For a volumetric flow rate of 1 ml h^{-1} through the device (Fig. 2a), the expected maximum velocity was 460 μm s^{-1} and occurred halfway between microposts (Supplementary Fig. 1c). We also observed that the anticipated maximum shear stress experienced by a cell near the surface of a micropost is 0.4 dyn cm^{-2} at $\theta = 68°$ (Supplementary Fig. 1d), well below the levels that are physiologically deleterious and within the range at which maximum cell attachment would be expected to occur, as determined experimentally by linear shear stress chamber studies (Supplementary Fig. 2). In summary, the analysis indicated an equilateral triangular micropost arrangement with a 50 μm distance between microposts and with a 50 μm shift after every 3 rows of microposts to be the most efficient micropost geometric arrangement and spacing. Applying a volumetric flow rate of 1 ml h^{-1} through the device yields high throughput efficient capture of cells with low non-specific binding

Making the surface functional. The CTC-chip surface was made functional with EpCAM antibodies using avidin–biotin chemistry[32]. The surface of the chip was

modified with 4% (v/v) 3-mercaptopropyl trimethoxysilane in ethanol at room temperature for 45 min, then treated with the coupling agent *N*-γ-maleimidobutyryloxy succinimide ester (GMBS, 1 μM), resulting in GMBS attachment to the microposts. Next, the chip was treated with 10 μg ml^{-1} of neutravidin at room temperature for 30 min, leading to immobilization onto GMBS, and then flushed with PBS to remove excess avidin. Finally, biotinylated EpCAM antibody at a concentration of 10 μg ml^{-1} in phosphate buffered solution (PBS) with 1% (w/v) BSA and 0.09% (w/v) sodium azide was allowed to react for 15–30 min before washing with PBS. The chip was air dried and stored at ambient temperature for up to three weeks until use.

Cell-line experiments. The human non-small-cell lung cancer (NSCLC) cell line NCI-H1650 was maintained and grown to confluence in RPMI-1640 medium containing 1.5 mM L-glutamine supplemented with 10% fetal bovine serum at 37 °C in 5% CO_2, with humidity. Growth medium was aspirated and cells incubated with trypsin for 10 min. A protein buffer was added to quench protease activity. Cells were then pre-labelled with cell tracker orange using the standard protocol provided by the manufacturer. The cell titre was determined by counting with a haemocytometer. The desired concentration of cells was then prepared by serial dilution of the original cell suspension in PBS. Labelled cells were spiked into whole blood.

Fixation and staining of captured cells. Captured cells were fixed by flowing 0.9 ml of 1% PFA in PBS, through the device at 3.0 ml h^{-1} for 20 min. The chip was subsequently washed with a solution of 0.9 ml of 0.2% Triton X-100 in PBS for 10 min to induce cellular permeability and allow for intracellular staining. To identify any bound lymphocytes, 0.9 ml of anti-CD45 stock solution (50 μl of antibody stock solution in 1 ml of PBS) was passed through the chip at 3 ml h^{-1} for 30 min, followed by a PBS wash to remove excess antibody. To identify epithelial cells, 0.9 ml of anti-cytokeratin stock solution (50 μl of antibody stock solution in 1 ml of PBS) was passed through the chip at 3 ml h^{-1} for 30 min, followed by a PBS wash. Finally, to permit the identification of cellular nuclei, 0.9 ml of DAPI solution (10 μl of DAPI reagent in 1 ml of deionized water) was passed through the chip at 3 ml h^{-1}, for 15 min, followed by a PBS wash. The chip was removed from the manifold, wiped dry near the fluid ports and stored in the dark at 4 °C until imaging.

Shear stress studies using linear shear Hele-Shaw chambers. Shear stress has an important role in cell capture. An optimum shear stress should be applied such that one can capture the maximum number of cancer cells at high enough flow rates. To find the optimal flow rate, we studied the effect of shear stress on the cell capture using Hele-Shaw microfluidic chambers[32]. The geometry of these chambers (Supplementary Fig. 2d) is such that the shear stress varies linearly along the chamber length (Supplementary Fig. 2e), permitting the study of a wide range of shear stresses for a given flow rate. Cultured lung cancer cells were spiked into PBS solution, and then passed through the Hele-Shaw chambers made functional with EpCAM antibody at a constant flow rate. As the shear stress decreased along the channel, the density of the cells adhering to the micropost surface increased (Supplementary Fig. 2a–c). The effect of shear stress on cell adhesion through EpCAM antibody–antigen binding (Supplementary Fig. 2f) indicated that 8 dyn cm^{-2} was the optimum shear rate, resulting in the capture of 200 cells mm^{-2} of functionalized capture surface.

Testing CTC capture from patient blood. Before the more extensive clinical testing described in the main text, we performed initial clinical experiments using blood samples from patients with advanced NSCLC. First, we evaluated CTC capture efficiency among whole blood samples and red blood cell lysed samples in six NSCLC patients. Isolated cells are shown in Supplementary Fig. 3a. The white arrows point to CTCs and the green arrows to leukocytes. The inset at the top left of shows a high-magnification view of a cytokeratin positive (cytokeratin^{+}) CTC and the inset at the lower left presents a CD45^{+} leukocyte. The total number of cells captured for each case was analysed to calculate captured CTCs per mL of blood (Supplementary Fig. 3b). Consistent with the spiked cell experiments, we observed no significant difference in the CTC counts of cancer patients from whole blood or lysed blood, confirming that the CTC-chip can be used to isolate CTCs directly from whole blood without any need for pre-processing. The reproducibility of CTC measurements between split samples was tested with a Wilcoxon matched-pairs signed-ranks two-sided test and found to be low.

Materials. 3-Mercaptopropyl trimethoxysilane was obtained from Gelest. Ethanol (200 proof), tissue culture flasks, a haemocytometer and serological pipettes were purchased from Fisher Scientific. Fetal bovine serum (FBS) and 0.5 M ethylene diamine tetra acetic acid (EDTA) were purchased from Gibco. Dimethyl sulfoxide (DMSO), sodium azide, lyophilized bovine serum albumin (BSA) and a glovebag for handling the moisture-sensitive silane were obtained from Sigma. The coupling agent GMBS (*N*-γ-maleimidobutyryloxy succinimide ester), NHS-LC-LC-biotin (succinimidyl-6¢-(biotinamido)-6-hexanamido hexanoate), and fluorescein-conjugated neutravidin were obtained from Pierce Biotechnology. Biotinylated mouse anti-human anti-EpCAM was obtained from R&D Systems. Human non-small-cell lung cancer line NCI-H1650, prostate cell line PC3-9, breast cancer cell line SKBr-3 and bladder cancer cell line T-24 were purchased from American Type Culture Collection, and RPMI-1640 growth medium was purchased from Invitrogen. Orange (5- (and 6-)-(((4-chloromethyl)-benzoyl) amino) tetramethyl-rhodamine; CMTMR) and green (5-chloromethylfluorescein diacetate; CMFDA) cell tracker dyes were obtained from Molecular Probes. Anti-cytokeratin PE (CAM 5.2, conjugated with phycoerythrin), CD45 FITC, the fluorescent nucleic acid dye nuclear dye 4′,6-diamidino-2-phenylindole (DAPI) and 10 ml vacutainer tubes were purchased from BD Biosciences.

31. Drummond, J. E. & Tahir, M. I. Laminar viscous flow through regular arrays of parallel solid cylinders. *Int. J. Multiphase Flow* 10, 515–539 (1984).
32. Murthy, S. K., Sin, A., Tompkins, R. G. & Toner, M. Effect of flow and surface conditions on human lymphocyte isolation using microfluidic chambers. *Langmuir* 20, 11649–11655 (2004).

LETTERS

Performance variability enables adaptive plasticity of 'crystallized' adult birdsong

Evren C. Tumer[1] & Michael S. Brainard[1,2]

Significant trial-by-trial variation persists even in the most practiced skills. One prevalent view is that such variation is simply 'noise' that the nervous system is unable to control or that remains below threshold for behavioural relevance[1–3]. An alternative hypothesis is that such variation enables trial-and-error learning, in which the motor system generates variation and differentially retains behaviours that give rise to better outcomes. Here we test the latter possibility for adult bengalese finch song. Adult birdsong is a complex, learned motor skill that is produced in a highly stereotyped fashion from one rendition to the next[4,5]. Nevertheless, there is subtle trial-by-trial variation even in stable, 'crystallized' adult song[6–8]. We used a computerized system to monitor small natural variations in the pitch of targeted song elements and deliver real-time auditory disruption to a subset of those variations. Birds rapidly shifted the pitch of their vocalizations in an adaptive fashion to avoid disruption. These vocal changes were precisely restricted to the targeted features of song. Hence, birds were able to learn effectively by associating small variations in their vocal behaviour with differential outcomes. Such a process could help to maintain stable, learned song despite changes to the vocal control system arising from ageing or injury. More generally, our results suggest that residual variability in well learned skills is not entirely noise but rather reflects meaningful motor exploration that can support continuous learning and optimization of performance.

Birdsong, like speech, is a learned motor skill that requires exceptionally precise and rapid control of vocal musculature; for both song and speech, modulation in acoustic structure occurs on a timescale of milliseconds[4,9]. This fine level of vocal control is learned in a process that depends crucially on auditory feedback. In many songbird species the learned song 'crystallizes' into a stable form after reaching adulthood[4]. However, disruption of feedback can cause a gradual deterioration in song, suggesting the continued importance of feedback for monitoring and maintaining performance[10–14].

Prior modelling indicates that song learning could proceed through trial and error, in which variable songs are produced ('motor exploration') followed by selective reinforcement of better versus worse variants[15–18]. Consistent with the possibility that song variation reflects motor exploration, experimental data indicate that a component of variability in both juvenile and adult song is centrally controlled[7,8,19–22]. Despite these observations, it remains unknown whether birds can modify their songs by a trial-and-error process in which they associate variations in vocal behaviour with better versus worse outcomes.

We directly tested this possibility in adult bengalese finches by using a computerized system to continuously monitor song and provide online disruptive feedback for a targeted range of vocalizations. A normal song consists of a stereotyped sequence of discrete 'syllables' (Fig. 1a). We focused our experiments on the pitch, or fundamental frequency, of individual syllables, because fundamental frequency is learned, precisely controlled and readily quantifiable[5,7,8]. The coefficient of variation of fundamental frequency for syllables in our study was 1.5% (median; range 1.2–2%). This only slightly exceeds the variation present across renditions of notes by trained vocalists and is much less than the variation across repeated phonemes for normal speech[23,24]. Moreover, this variation is of the same order as frequency discrimination thresholds for birds and humans[25]. Hence, variation in adult bengalese finch song is quite subtle, commensurate with that of other well-practiced skills.

An experiment to assess whether this variability can be used by the nervous system is illustrated in Fig. 1. We targeted fundamental frequency of syllable 'a' (Fig. 1a) for modification. Initially, fundamental frequency was tightly controlled with only ~1% variation (Fig. 1c; fundamental frequency = 2,281 ± 27 Hz). We used short bursts of white noise as a negative reinforcer for a subset of syllable variants, as white noise disrupts normal feedback and causes a gradual, non-specific deterioration of song when chronically used to disrupt feedback (see Methods). Here, rather than indiscriminately disrupting feedback, we delivered white noise in a manner contingent on acoustic features of the targeted syllable. Higher pitched renditions received a short latency burst of white noise ('hit'), whereas lower pitched versions received normal feedback ('escape') (Fig. 1b). Unlike natural feedback, which potentially provides fine-grained information about amplitude, pitch and other features of song, our experimentally controlled feedback provided a binary signal for each rendition of the target syllable, either 'bad' (white noise burst) or 'good' (escape), without further detail about the aspects of song that triggered feedback. We then asked whether birds could detect the externally imposed contingency that led to disruption of feedback and alter their songs to escape it.

This form of behaviourally contingent reinforcement drove rapid, adaptive changes to targeted features of adult song. For the experiment of Fig. 1, the fundamental frequency of the targeted syllable decreased significantly over the first three days of reinforcement (Fig. 1d; fundamental frequency reduced by 1.8 times the standard deviation of the baseline distribution; $P < 0.0001$). Correspondingly, the hit rate dropped from 91% (baseline) to 17%. By establishing the appropriate contingency, the same reinforcement scheme could direct either increases or decreases in pitch (Fig. 1e). When white noise was differentially delivered at higher frequencies, the fundamental frequency always shifted lower; when white noise bursts were delivered at lower frequencies, the fundamental frequency shifted higher (Fig. 1f). Hence, the applied contingency determined the direction of behavioural change ($P < 0.01$; sign-test for 7 of 7 cases shifting in the direction of the applied contingency). In all cases, changes in fundamental frequency led to a large reduction in hit rate,

[1]Keck Center for Integrative Neuroscience, Sloan-Swartz Center for Theoretical Neurobiology, Department of Physiology, University of California, San Francisco, California 94143-0444, USA. [2]Department of Psychiatry, University of California, San Francisco, California 94143, USA.

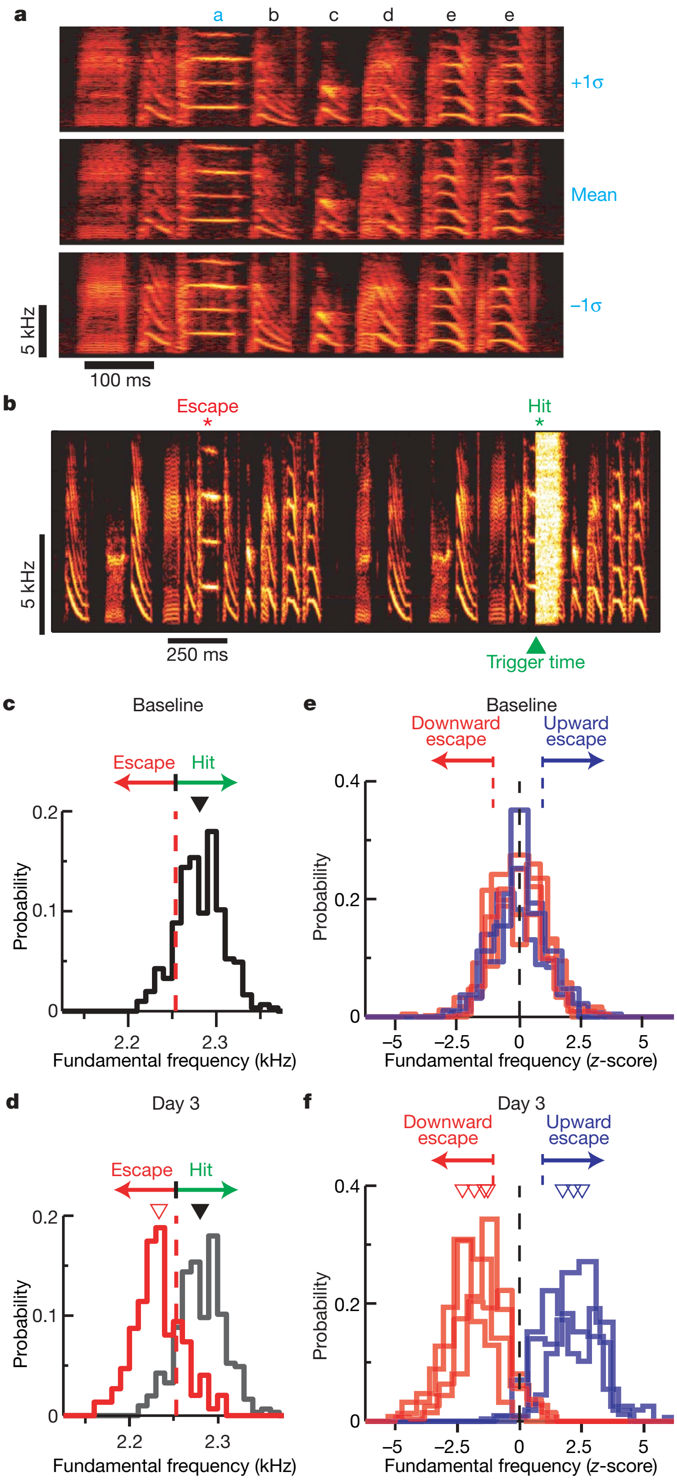


Figure 1 | Differential reinforcement can adaptively alter features of adult song. a, Syllables (a, b, c, d, e, e) are normally produced with little variation. Three songs are shown for which the fundamental frequency of 'a' spanned 2 standard deviations of the baseline distribution (Supplementary Recording 1 contains corresponding audio files). **b**, White noise bursts ('hits') were targeted at higher pitched versions of 'a'. **c**, Baseline fundamental frequency distribution for 'a', showing overall mean (triangle) and mean for escapes (line). **d**, Fundamental frequency distribution after 3 days white noise. **e**, **f**, Baseline (**e**) and day 3 (**f**) distributions for seven experiments in which white noise directed either downward (red) or upward (blue) shifts in fundamental frequency. Fundamental frequency is expressed in units of the standard deviation of the baseline distribution (*z*-score, see Methods).

with the median decreasing from 91% to 21% by the third day. These data provide the first evidence that auditory reinforcement signals can direct specific, adaptive changes to adult song. As a corollary, they demonstrate that birds were able to effectively associate small natural variations in vocal production with differential outcomes.

Changes elicited by differential reinforcement occurred rapidly. A detailed examination of one experiment reveals that fundamental frequency approached the asymptotic range within one day (Fig. 2a). Similarly, for six of seven birds, significant adaptive changes to fundamental frequency occurred within 7 h (Fig. 2b, 'morning') and for all birds significant changes occurred within the first day (Fig. 2b, 'evening'). The median number of syllables sung within the first half-day and full day were 605 and 1,179, respectively. By day 3, fundamental frequency stabilized at nearly asymptotic values (Fig. 2b).

The induced changes in fundamental frequency recovered rapidly. After at least 3 days of reinforcement, contingent white noise bursts were terminated. In every case, the fundamental frequency reverted towards its original range (Fig. 2c). Hence, the nervous system retains a representation of the initial song and both the capacity and impetus to return song towards its original structure in the absence of externally imposed drive.

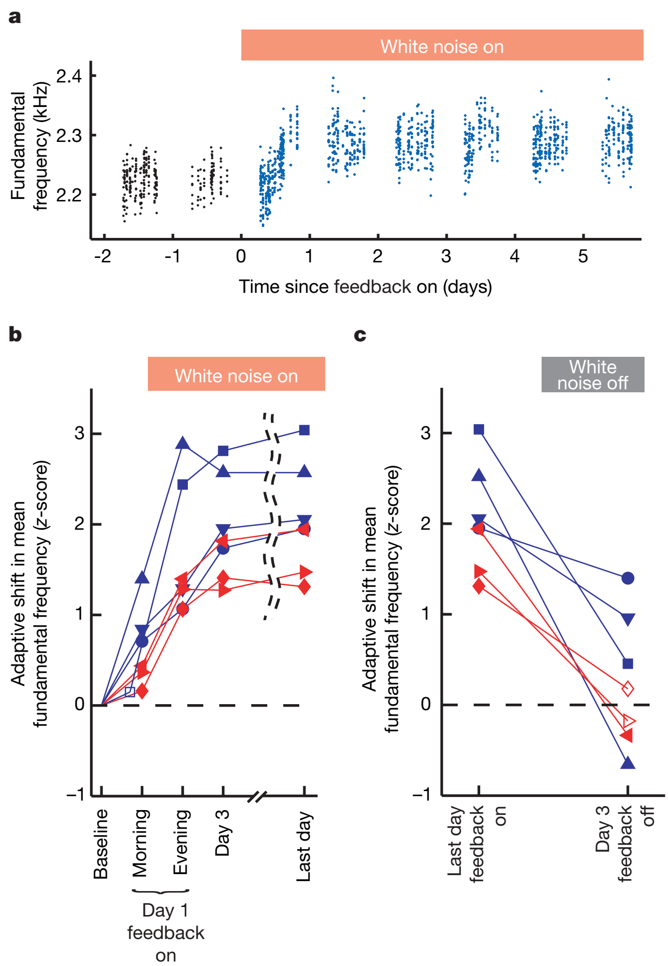


Figure 2 | Adaptive shifts in fundamental frequency occur rapidly and recover. a, Fundamental frequency of targeted syllables for one adult bird (age 334 days). Fundamental frequency progressively increased during the first day. **b**, Upward (blue) and downward (red) shifts in fundamental frequency. Filled symbols indicate significance ($P < 0.05$). After 3 days, fundamental frequency changed little, indicating that shifts had reached near asymptotic values. **c**, Fundamental frequency for last day with white noise feedback on ('Last day, feedback on') and third day following termination of feedback ('Day 3, feedback off').

In principle, birds could escape white noise equally well by altering the fundamental frequency of only the targeted syllable or of larger segments of song. Nevertheless, we consistently found that changes to fundamental frequency were restricted exclusively to targeted syllables (Fig. 3a, b, red symbols). Other syllables, even when they occurred within a few tens of milliseconds, did not change (Fig. 3a, b, blue symbols). Furthermore, the changes were specific to fundamental frequency and did not affect other features such as duration, volume and spectral entropy (Fig. 3c). Hence, even though the experimentally imposed contingency between performance and feedback was not revealed to the bird in any direct sense, the nervous system was able to detect and respond precisely to that contingency.

This specificity indicates an impressive capacity of the nervous system to modify discrete features of song independently. This is appropriate for vocal learning, where birds match their song to rapidly varying features of an acoustic model. If modification of one song feature generalized to cause modification of others, learning might still proceed, but such interference would probably slow its progress[26].

In theory, reinforcement signals can drive learning even at long latencies to the actions that precipitate them[15]. For complex motor skills, however, the nervous system might detect the contingency more effectively at shorter delays. We tested the importance of delay by varying the time between measurement of fundamental frequency

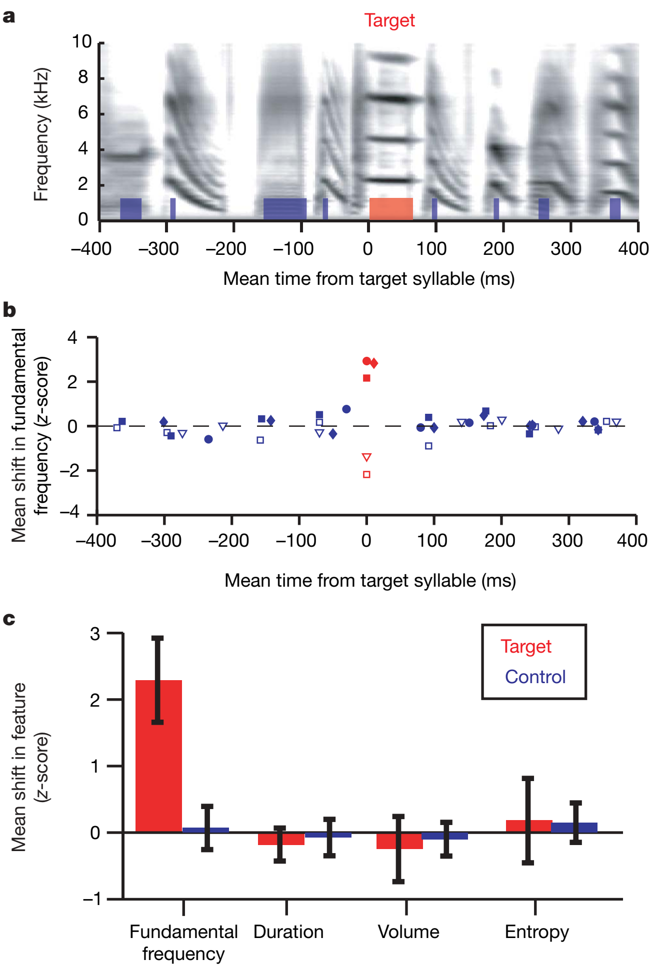


Figure 3 | Changes are restricted to targeted features of song. a, Spectrogram illustrating analysed features for target (red bar) and control (blue bars) syllables of an individual experiment. **b,** Mean changes in fundamental frequency for target (red) and control (blue) syllables. Squares represent two experiments for song illustrated in **a**. Data from 3 additional birds (circles, triangles, diamonds) are shown without corresponding spectrograms. Filled and open symbols indicate experiments with upward and downward shifts in fundamental frequency, respectively. **c,** Spectral characteristics other than fundamental frequency were not altered for either target (red) or control (blue) syllables. Bars indicate mean ± standard deviation.

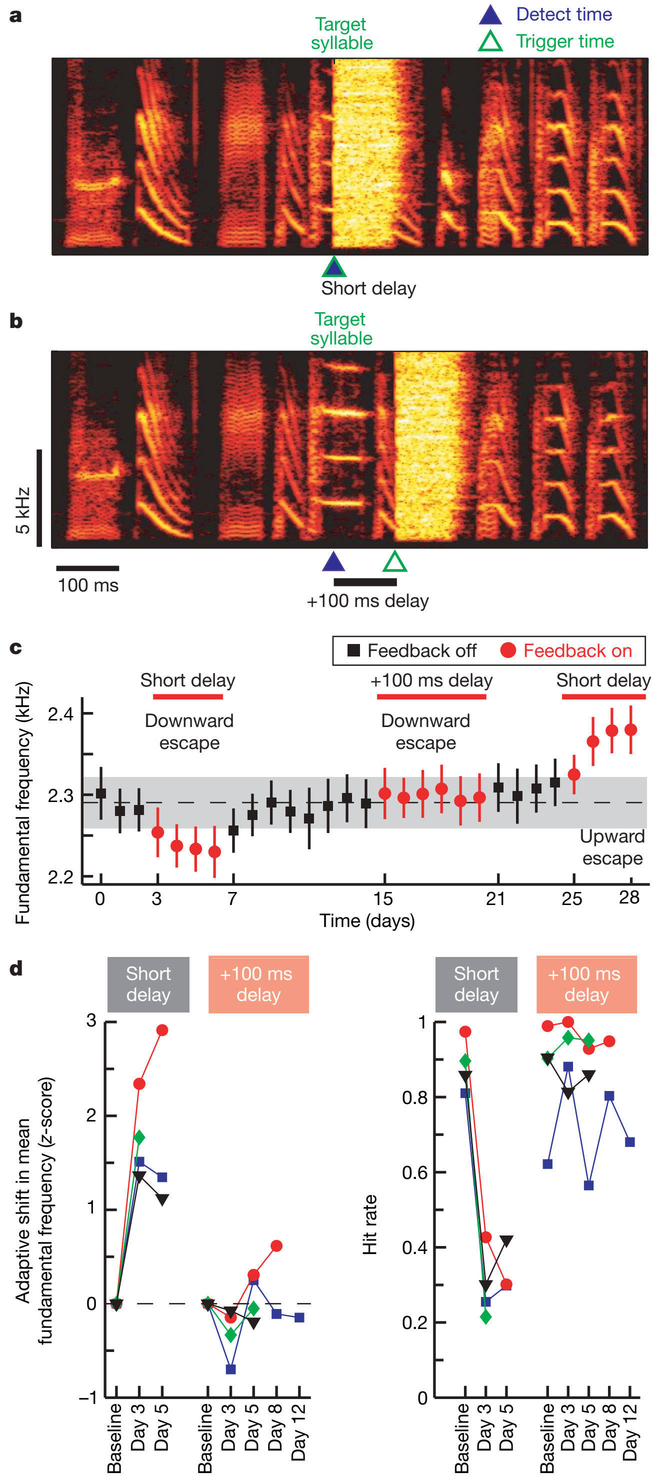


Figure 4 | Delayed feedback prevents adaptive pitch shifts. a, b, Spectrograms illustrating short delay (**a**) and +100 ms delay (**b**) feedback for one bird. **c,** Fundamental frequency (mean ± standard deviation) for this bird. Initially, short delay feedback directed a downward shift in fundamental frequency. After recovery, the same contingency with delayed feedback was ineffective. Subsequently, an upward shift was driven using a new short delay contingency. Shading indicates baseline ±1 standard deviation. **d,** Summary of shifts in fundamental frequency and hit rate. Symbols indicate 4 birds subjected to both short delay and +100 ms delay reinforcement. Shifts were prevented when feedback was delayed.

and delivery of feedback. White noise typically was delivered within 16–30 ms. In this case, white noise overlapped with the targeted syllable (Fig. 4a, 'short delay'). For some experiments, white noise was delayed by an extra 100 ms. In this case, white noise started after completion of the targeted syllable (Fig. 4b, '+100 ms delay'). Syllables targeted at short delay rapidly shifted and reached asymptotic values after roughly 3–4 days. In contrast, the same syllables, when targeted with delayed feedback, exhibited no directed change to fundamental frequency even over longer periods (Fig. 4c, d). Hence, even small delays profoundly attenuated adaptive responses. Likewise, in the +100 ms condition, there were no systematic changes to syllables following the targeted syllable, for which white noise was delivered in a non-contingent manner (see Supplementary Fig. 3). Hence, rapid changes to fundamental frequency required both that white noise was contingent on the structure of a syllable and that it occur at short delay. These data suggest that the vocal control system may take advantage of normally predictable timing between premotor activity and resultant sensory consequences; normally, activity of premotor neurons influences song and resultant feedback at latencies < 70 ms (ref. 27). Therefore, the nervous system may not readily detect reinforcement contingencies when feedback is delayed beyond this range. Although such timing constraints are not necessary for reinforcement learning, they potentially enable more efficient learning and may be a general feature of sensorimotor systems where there is predictable delay between action and consequence[28].

Our results indicate that small natural variations present in adult song can enable adaptive modification of targeted features. These changes exceed the frequency discrimination threshold for birds[25]. Moreover, the reversion of fundamental frequency after termination of feedback (Fig. 2c) indicates the nervous system is sensitive to the discrepancy from the original song. Nevertheless, our reinforcement scheme constrained shifts in fundamental frequency to be similar in magnitude to the initial range of variation (see Supplementary Fig. 2). This raises the question of whether changes in adult song are restricted to subtle fine-tuning, or whether more dramatic remodelling is possible. We tested this by progressively altering the reinforcement contingency so that continued escape from white noise would require progressively larger shifts in fundamental frequency (Fig. 5). We were able to incrementally drive large changes in fundamental frequency, such that syllables eventually were produced in a range completely non-overlapping with the baseline range (Supplementary Movie 1 demonstrates the salience of these changes). This finding implies that following each incremental shift in fundamental frequency a new range of behaviourally relevant variation was established that enabled differential reinforcement of more extreme vocalizations. The range of variation remained relatively constant over time (for example, see Figs 1, 4, 5). Hence, as learning progresses, current performance is continuously surrounded by a 'halo' of variation that enables continuous adaptive modification.

In summary, our results demonstrate that binary feedback signals can drive rapid plasticity of normally stable adult song. In contrast to previous studies in which disruption of feedback led to a deterioration of song[10–14], we show that differentially delivered feedback can direct precise, adaptive changes to adult song. The changes are precise in that they are restricted to targeted features. They are adaptive in that birds alter those features systematically to escape disrupted feedback. These data provide empirical support for models suggesting that song might normally be shaped by reinforcement signals[16–18]. More generally, they indicate that the trial-and-error process of reinforcement learning can efficiently guide adaptive modification of highly complex and tightly controlled motor skills.

The adaptive plasticity reported here also demonstrates that birds can accurately associate small variations in fundamental frequency with resultant consequences. In principle, variation could be monitored by sensory feedback (auditory or proprioceptive). Alternatively, if variation is centrally generated, it could be monitored internally by an efference copy of premotor activity[17]. Consistent with this possibility, studies in both birds and mammals indicate that some movement variability derives from central neural activity rather than the periphery[7,8,29,30], suggesting that variability may be actively generated for motor exploration. Regardless of mechanism, our data indicate that natural variations present in crystallized adult song are not simply noise but rather can be exploited for trial-and-error

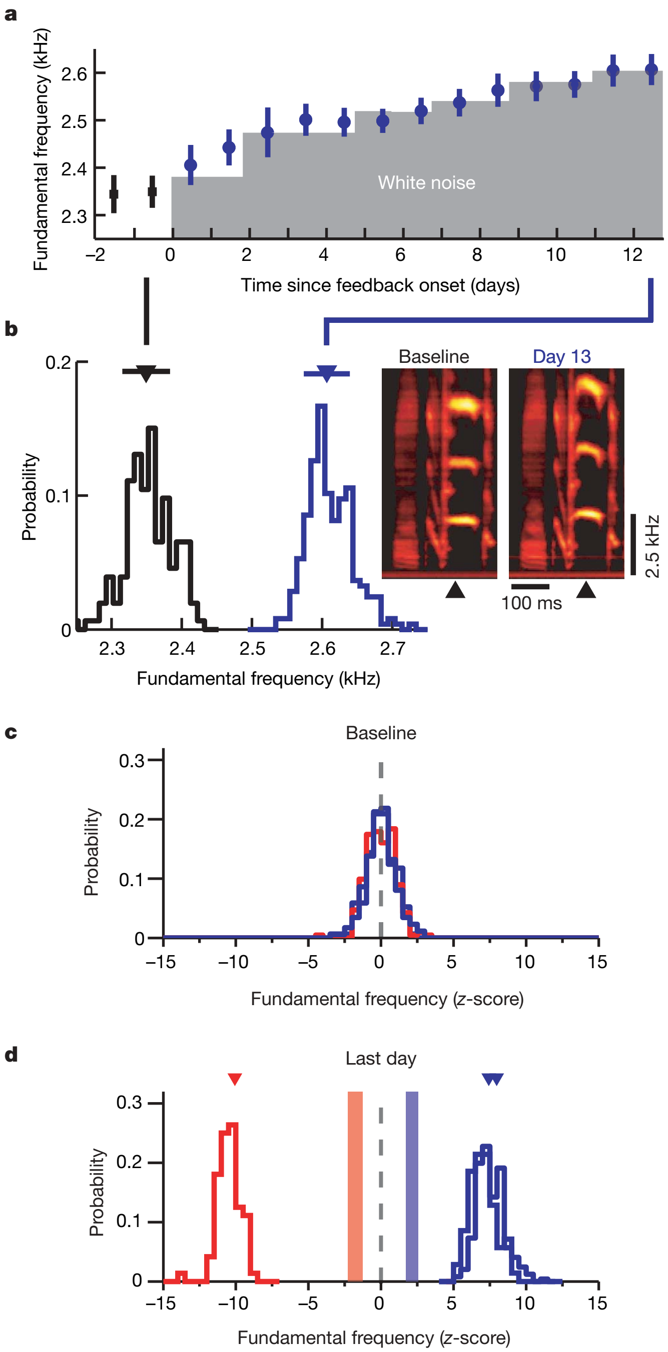


Figure 5 | Incremental adjustment of threshold drives large pitch changes. **a**, Changes to fundamental frequency for one experiment. Points indicate mean (± standard deviation) on each day. Shading indicates threshold for escapes. **b**, Fundamental frequency distributions for baseline (black) and day 13 (blue). Spectrograms show the average syllable for corresponding days. **c**, **d**, Baseline (**c**) and final (**d**) fundamental frequency distributions for three experiments with thresholds incrementally adjusted upward (blue) or downward (red). Over 11–14 days, mean fundamental frequency (triangles) shifted by 7.0–10.5 standard deviations of the baseline distribution (226–376 Hz). For comparison, vertical shaded regions show the range of mean changes driven with a fixed reinforcement contingency.

learning. We suggest that in other systems as well, subtle variation in performance may reflect continued experimentation to optimize behaviour, rather than noise that the nervous system is unable to control, or that is below threshold for behavioural relevance.

METHODS SUMMARY

Subjects. Eleven adult (age 112–346 days) male bengalese finches (*Lonchura striata var. domestica*) were used. All procedures were approved by the University of California, San Francisco Institutional Animal Care and Use Committee.

Song recording and behaviourally contingent reinforcement. Birds were housed individually in sound isolation chambers throughout the experiments. Custom LabView software (National Instruments) was used for song recording and online detection of target syllables. We tuned spectral templates to differentially detect and trigger reinforcement of higher or lower pitched syllables, thereby establishing a clear direction in which pitch would need to be shifted to reduce disruptive feedback.

Hit rate and syllable structure calculations. Hit rate was defined as the percentage of times the target syllable matched the established criterion for white noise playback out of the total number of target syllables sung. Fundamental frequency was calculated from the frequency bins surrounding the harmonic peaks in the power spectrum of the target syllable. Volume of individual syllables was normalized by mean volume of the entire song. Duration and spectral entropy (from 0.5–10 kHz) were calculated as previously described[27]. Significance was determined using two-tailed *t*-tests. During baseline periods, hit rate was calculated in the absence of white noise playback. During the period of contingent reinforcement, we randomly selected a subset of songs as 'catch' trials for which white noise playbacks were suppressed (10–20%). The values reported for hit rate, fundamental frequency and other syllable parameters were derived from these catch trials.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 August; accepted 16 October 2007.

1. Beers, R. J. V., Baraduc, P. & Wolpert, D. M. Role of uncertainty in sensorimotor control. *Phil. Trans. R. Soc. Lond. B* **357,** 1137–1145 (2002).
2. Todorov, E. Optimality principles in sensorimotor control. *Nature Neurosci.* **7,** 907–915 (2004).
3. Harris, C. M. & Wolpert, D. M. Signal-dependent noise determines motor planning. *Nature* **394,** 780–784 (1998).
4. Konishi, M. Birdsong: From behavior to neuron. *Annu. Rev. Neurosci.* **8,** 125–170 (1985).
5. Deregnaucourt, S. *et al.* How sleep affects the developmental learning of bird song. *Nature* **433,** 710–716 (2005).
6. Sossinka, R. & Bohner, J. Song types in the zebra finch *Poephila guttata castanotis*. *Z. Tierpsychol.* **53,** 123–132 (1980).
7. Kao, M. H., Doupe, A. J. & Brainard, M. S. Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. *Nature* **433,** 638–643 (2005).
8. Kao, M. H. & Brainard, M. S. Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. *J. Neurophysiol.* **96,** 1441–1455 (2006).
9. Doupe, A. J. & Kuhl, P. K. Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* **22,** 567–631 (1999).
10. Nordeen, K. W. & Nordeen, E. J. Auditory feedback is necessary for the maintenance of stereotyped song in adult zebra finches. *Behav. Neural Biol.* **57,** 58–66 (1992).
11. Okanoya, K. & Yamaguchi, A. Adult Bengalese finches (*Lonchura striata* var. *domestica*) require real-time auditory feedback to produce normal song syntax. *J. Neurobiol.* **33,** 343–356 (1997).
12. Woolley, S. M. & Rubel, E. W. Bengalese finches *Lonchura Striata domestica* depend upon auditory feedback for the maintenance of adult song. *J. Neurosci.* **17,** 6380–6390 (1997).
13. Leonardo, A. & Konishi, M. Decrystallization of adult birdsong by perturbation of auditory feedback. *Nature* **399,** 466–470 (1999).
14. Zevin, J. D., Seidenberg, M. S. & Bottjer, S. W. Limits on reacquisition of song in adult zebra finches exposed to white noise. *J. Neurosci.* **24,** 5849–5862 (2004).
15. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, Massachusetts, 1998).
16. Doya, K. & Sejnowski, T. in *The New Cognitive Neurosciences* (ed. Gazzaniga, M.) 469–482 (MIT Press, Cambridge, Massachusetts, 2000).
17. Troyer, T. & Doupe, A. J. An associational model of birdsong sensorimotor learning. I. Efference copy and the learning of song syllables. *J. Neurophys.* **84,** 1204–1223 (2000).
18. Fiete, I. R., Fee, M. S. & Seung, H. S. Model of birdsong learning based on gradient estimation by dynamic perturbation of neural conductances. *J. Neurophys.* **98,** 2038–2057 (2007).
19. Olveczky, B. P., Andalman, A. S. & Fee, M. S. Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *PLoS Biol.* **3,** e153 (2005).
20. Hessler, N. A. & Doupe, A. J. Social context modulates singing-related neural activity in the songbird forebrain. *Nature Neurosci.* **2,** 209–211 (1999).
21. Bottjer, S. W., Miesner, E. A. & Arnold, A. P. Forebrain lesions disrupt development but not maintenance of song in passerine birds. *Science* **224,** 901–903 (1984).
22. Scharff, C. & Nottebohm, F. A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. *J. Neurosci.* **11,** 2896–2913 (1991).
23. Amir, O., Amir, N. & Kishon-Rabin, L. The effect of superior auditory skills on vocal accuracy. *J. Acoust. Soc. Am.* **113,** 1102–1108 (2003).
24. Sundberg, J., Prame, E. & Iwarsson, J. in *Vocal Fold Physiology, Controlling Complexity and Chaos* (eds Davis, P. & Fletcher, N.) 291–306 (Singular Publishing Group, San Diego, 1996).
25. Dent, M. L., Dooling, R. J. & Pierce, A. S. Frequency discrimination in budgerigars (*Melopsittacus undulatus*): Effects of tone duration and tonal context. *J. Acoust. Soc. Am.* **107,** 2657–2664 (2000).
26. Fiete, I. R., Hahnloser, R. H., Fee, M. S. & Seung, H. S. Temporal sparseness of the premotor drive is important for rapid learning in a neural network model of birdsong. *J. Neurophysiol.* **92,** 2274–2282 (2004).
27. Sakata, J. T. & Brainard, M. S. Real-time contributions of auditory feedback to avian vocal motor control. *J. Neurosci.* **26,** 9619–9628 (2006).
28. Held, R., Efstathiou, A. & Greene, M. Adaptation to displaced and delayed visual feedback from the hand. *J. Exp. Psychol.* **72,** 887–891 (1966).
29. Churchland, M. M., Afshar, A. & Shenoy, K. V. A central source of movement variability. *Neuron* **52,** 1085–1096 (2006).
30. Churchland, M. M., Santhanam, G. & Shenoy, K. V. Preparatory activity in premotor and motor cortex reflects the speed of the upcoming reach. *J. Neurophysiol.* **96,** 3130–3146 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Doupe, L. Frank, T. Warren, M. Wohlgemuth, S. Sober, J. Sakata, C. Hampton and J. Wong for comments. This work was supported by an NIDCD NRSA postdoctoral fellowship and the Sloan-Swartz Foundation (E.C.T.) and by an NIDCD R01 award, an NIMH Conte Center for Neuroscience Research award and a McKnight Foundation Scholars Award (M.S.B).

Author Contributions E.C.T. performed the experiments and analysis; E.C.T. and M.S.B. designed the experiments and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to evren@phy.ucsf.edu.

METHODS

Song recording and behaviourally contingent reinforcement. Birds were housed individually in sound isolation chambers (Acoustic Systems) with a 14/10 h light/dark cycle throughout the experiments. Custom LabView software (National Instruments) was used for song recording and online detection of target syllables. Before initiation of contingent reinforcement, songs were recorded for several days to measure baseline statistics on fundamental frequency of targeted syllables and develop spectral templates to detect syllables in real-time and trigger playback of white noise contingent on syllable spectral structure. For 2 birds a diurnal trend in pitch was apparent during baseline, reminiscent of that in juvenile zebra finches[5] (for example, Fig. 2a), but did not correlate with observed plasticity. Templates were constructed from the power spectrum of short segments (8 ms) of the target syllable. By varying the exemplar syllables from which templates were constructed, we tuned templates to differentially detect and trigger reinforcement of higher or lower pitched syllables. During the baseline period (with no white noise playback), we monitored the distribution of fundamental frequency for 'hits' and 'escapes' to determine empirically the actual feedback contingency that was applied in each experiment. For most experiments, 'escapes' were present only at the upper (or lower) end of the natural range of variation for fundamental frequency, establishing a clear direction in which pitch would need to be shifted to reduce disruptive feedback. For these experiments, an adaptive response was defined as a shift in mean pitch towards the end of the baseline distribution that contained escapes. See Supplementary Information for additional details on the timing and specificity of contingent reinforcement. For two experiments, escapes were present at both the upper and lower ends of the baseline distribution. In both of these cases, birds shifted the pitch of the targeted syllables upwards and reduced the hit rate. However, in these cases, the direction of adaptive response (increase versus decrease in pitch) was not clearly defined by the initial feedback contingency, and therefore these experiments are not included in analyses documenting that changes to targeted features of song can be specifically directed by predetermined feedback contingencies (for example, Fig. 1).

White noise bursts were used to differentially reinforce variations within the normal range of vocalizations. The duration of white noise bursts was long enough to overlap with the remainder of the targeted syllable (80–200 ms). We chose white noise bursts as our stimulus for differential reinforcement for two reasons. First, the superposition of white noise on normal feedback disrupts the sound that the bird experiences, and previous experiments have shown that chronic (non-contingent) disruption of feedback in this fashion results in a gradual and non-specific deterioration of song[14,31,32]. Second, perturbation of feedback with similar stimuli for both birds and humans can cause acute disruptions of ongoing vocalizations[27]. We confirmed for each of our experiments that white noise bursts were behaviourally salient in that they caused a significant increase in the probability of song stoppages during the first day of exposure. These acute effects typically habituated over the course of extended exposure. Nevertheless, they suggest that white noise stimuli superimposed on song are readily detected by the nervous system. Our finding that birds consistently alter their songs to avoid white noise bursts indicates that this form of feedback disruption indeed acts as an effective negative reinforcer.

Hit rate and fundamental frequency calculation. Hit rate was defined as the percentage of times the target syllable matched the established criterion for white noise playback out of the total number of target syllables the bird sang. During baseline periods, hit rate was calculated in the absence of white noise playback. During the period of contingent reinforcement, we additionally suppressed white noise playbacks for a randomly selected subset of songs (10–20%). These catch trials allowed us to monitor the progression of changes to pitch (and hit rate) for songs that were produced with entirely normal feedback, avoiding any potential complications due to online contributions of feedback to song structure. The values reported for fundamental frequency and hit rate are derived from these catch trials.

Online evaluation of fundamental frequency for purposes of contingent reinforcement relied on short (8 ms) segments of targeted syllables and a rapid template-matching algorithm (see above) in order to minimize the latency between production of song features that met criterion and the delivery of white noise. For analysis of spectral characteristics of targeted syllables, longer segments (16–64 ms) of the syllable waveform were used. Each segment was Fourier transformed and the fundamental frequency was calculated from the average of the five frequency bins neighbouring the peak weighted by the corresponding spectral power in each bin. Volume of individual syllables was normalized by mean volume of the entire song. Duration and spectral entropy (from 0.5–10 kHz) were calculated as previously described[27]. To summarize effects across syllables, we expressed changes in measured features as z-scores, computed as the difference between the means of the experimentally shifted and baseline distributions, divided by the standard deviation of the baseline distribution. Significance was determined using two-tailed t-tests.

31. Leonardo, A. Experimental test of the birdsong error-correction model. *Proc. Natl Acad. Sci. USA* **101**, 16935–16940 (2004).
32. Kozhevnikov, A. & Fee, M. S. Singing-related activity of identified HVC neurons in the zebra finch. *J. Neurophysiol.* **97**, 4271–4283 (2007).

Vol 450 | 20/27 December 2007 | doi:10.1038/nature05995

High-fidelity transmission of sensory information by single cerebellar mossy fibre boutons

Ede A. Rancz[1]*, Taro Ishikawa[1]*, Ian Duguid[1]*, Paul Chadderton[1]*, Séverine Mahon[1] & Michael Häusser[1]

Understanding the transmission of sensory information at individual synaptic connections requires knowledge of the properties of presynaptic terminals and their patterns of firing evoked by sensory stimuli. Such information has been difficult to obtain because of the small size and inaccessibility of nerve terminals in the central nervous system. Here we show, by making direct patch-clamp recordings *in vivo* from cerebellar mossy fibre boutons—the primary source of synaptic input to the cerebellar cortex[1,2]—that sensory stimulation can produce bursts of spikes in single boutons at very high instantaneous firing frequencies (more than 700 Hz). We show that the mossy fibre–granule cell synapse exhibits high-fidelity transmission at these frequencies, indicating that the rapid burst of excitatory postsynaptic currents underlying the sensory-evoked response of granule cells[3] can be driven by such a presynaptic spike burst. We also demonstrate that a single mossy fibre can trigger action potential bursts in granule cells *in vitro* when driven with *in vivo* firing patterns. These findings suggest that the relay from mossy fibre to granule cell can act in a 'detonator' fashion, such that a single presynaptic afferent may be sufficient to transmit the sensory message. This endows the cerebellar mossy fibre system with remarkable sensitivity and high fidelity in the transmission of sensory information.

We have used a parallel approach to make patch-clamp recordings from cerebellar mossy fibre boutons both *in vivo* and *in vitro*. In cerebellar slices, we targeted patch-clamp recordings to small structures (diameter about 2–3 μm) visible among the granule cells. Biocytin filling and subsequent processing (Fig. 1a; $n = 12$) confirmed their identity as mossy fibre boutons, revealing characteristic thorny excrescences on collaterals of a thin axon emerging from the white matter[1,2]. Mossy fibre boutons had a distinctive set of electrophysiological characteristics, including an absence of spontaneous synaptic potentials or currents, a small membrane capacitance ($C_m = 1.81 \pm 0.46$ pF (mean ± s.e.m.), $n = 7$; Supplementary Fig. 2), a high input resistance ($R_{input} = 520 \pm 88$ MΩ, $n = 12$), pronounced outward rectification, and time-dependent 'sag' on hyperpolarization (Fig. 1c). Using patch pipettes of similar tip sizes and geometries, we made recordings from mossy fibre boutons in the granule cell layer of cerebellar cortex *in vivo*, identified on the basis of their subthreshold and suprathreshold electrophysiological properties (Supplementary Table 1), which were identical with those observed *in vitro*. Mossy fibre boutons *in vivo* showed no detectable spontaneous synaptic events (Supplementary Fig. 3) and, when recovered, had a similar morphology (Fig. 1b) to those *in vitro*.

Mossy fibre boutons *in vivo* fired action potentials spontaneously at a rate of 3.9 ± 0.8 Hz ($n = 10$; Fig. 1d), distinct from the spontaneous firing rates of interneurons in the granule cell layer and of granule cells *in vivo*[3–5]. Strong hyperpolarization of mossy fibre boutons (below −90 mV) or voltage clamp (at −90 mV) did not prevent spontaneous firing ($n = 5$), suggesting that action potential initiation occurred electrotonically distant from the site of recording. However, action potential amplitude showed a marked dependence on membrane potential, decreasing with depolarization (Fig. 1g, h), consistent with local Na^+-channel inactivation limiting active invasion of the bouton at depolarized membrane potentials[6,7]. Previous extracellular recordings from putative mossy fibre units *in vivo* have reported firing rates of up to 1 kHz (refs 8–11). To determine the firing rates that single mossy fibre boutons can sustain, we injected brief current pulses to generate action potentials. Whereas long step-current injections produced only a single action potential (Fig. 1c, d), the injection of brief current pulses at high frequency resulted in repetitive firing at 500 Hz or greater (Fig. 1e, f; $n = 6$). These data indicate that cerebellar mossy fibres are capable of faithfully generating spikes at extremely high frequencies.

Spiking evoked by peripheral sensory stimulation was observed in both cell-attached and subsequent whole-cell recordings from mossy fibre boutons. Sensory stimulation reliably generated bursts of spikes (4.2 ± 0.6 action potentials with an average inter-spike interval (ISI) of 12.4 ± 1.7 ms; $n = 3$ responding boutons out of 10 boutons tested; Fig. 2a–c). The onset latency of spiking was 27.0 ± 19.2 ms, consistent with the onset latency of granule-cell excitatory postsynaptic currents (EPSCs) with the same stimulus and presumably reflecting a mixture of direct trigeminocerebellar and corticopontine input[12]. During the sensory-evoked presynaptic burst, the waveform of the action potential remained relatively constant, with the amplitude and maximal rate of rise decreasing by $15.3 \pm 10.3\%$ ($P > 0.1$) and $24.7 \pm 8.4\%$ ($P < 0.05$), respectively, within the burst, and with spike width increasing by only $4.9 \pm 2.7\%$ ($0.05 < P < 0.1$).

Sensory stimulation evokes a burst of EPSCs in cerebellar granule cells[3], the primary synaptic target of cerebellar mossy fibres. Because each granule cell receives only a small number of mossy fibre inputs, we determined whether transmission through a single mossy fibre could account for the postsynaptic burst (Supplementary Fig. 1) by comparing the dynamics of sensory-evoked presynaptic and postsynaptic responses. First, the number of sensory-evoked postsynaptic EPSCs (5.4 ± 0.7, $n = 14$; Fig. 2d, e) was indistinguishable from the number of sensory-evoked spikes in single presynaptic mossy fibre boutons ($P = 0.45$). Second, the latency of the first sensory-evoked presynaptic spike was similar to that of the first sensory-evoked EPSC (16.8 ± 1.9 ms, $P = 0.26$), and the timing of individual spikes was highly irregular (the coefficient of variation of evoked action potential intervals was 1.00 ± 0.06; Fig. 2c), comparable to the coefficient of variation of evoked postsynaptic EPSCs (0.69 ± 0.03, $n = 14$; Fig. 2d). Third, the minimum ISI of sensory-evoked action potentials was short (1.9 ± 0.5 ms, $n = 3$; Fig. 2b, e) and not significantly different from the minimum inter-event intervals (IEIs) of the postsynaptic EPSCs (3.1 ± 0.5 ms, $P = 0.27$; Fig. 2e). Finally, the mean ISI

[1]Wolfson Institute for Biomedical Research and Department of Physiology, University College London, Gower Street, London WC1E 6BT, UK.
*These authors contributed equally to this work.

of the presynaptic spikes was also not different from the mean IEI (13.5 ± 1.0 ms, $P = 0.64$; Fig. 2e). Because the number, pattern and frequency of sensory-evoked EPSCs were indistinguishable from those of the spikes in the sensory-evoked presynaptic spike burst, our findings are consistent with the idea that a single mossy fibre could provide the synaptic input observed in granule cells during sensory stimulation.

To further understand information transmission at the mossy fibre–granule cell synapse, we compared synaptic properties *in vivo* with those *in vitro*. We activated single mossy fibre–granule cell inputs using minimal stimulation in cerebellar slices[13] (Supplementary Fig. 4). Given the uncertainty about the extracellular calcium concentration ($[Ca^{2+}]_e$) *in vivo*[14,15] and the sensitivity of neurotransmitter release probability to $[Ca^{2+}]_e$, we compared EPSC amplitudes and paired-pulse ratios in the presence of 1.2 or 2 mM $[Ca^{2+}]_e$ with those recorded *in vivo* (Fig. 3). The average amplitudes and paired-pulse ratios of EPSCs *in vivo* fell between those recorded in 1.2 and 2 mM $[Ca^{2+}]_e$ *in vitro* (Fig. 3), suggesting that the release probability *in vivo* was between those obtained in these two *in vitro* conditions.

Given that the instantaneous frequency of presynaptic action potentials can exceed 700 Hz (Fig. 2b), we tested whether mossy fibre boutons can release transmitter at such high frequencies. Unitary mossy fibre EPSCs could reliably follow presynaptic stimulation at extremely high rates, up to 800 Hz for the first two stimuli and up to 500 Hz for the fifth stimulus (Fig. 4a, b). This suggests that single mossy fibre boutons can release transmitter at frequencies comparable to the maximal frequency of presynaptic action potentials and EPSCs observed *in vivo*. Similar to the result for paired-pulse depression (Fig. 3), synaptic depression during high-frequency trains in 1.2 and 2 mM $[Ca^{2+}]_e$ *in vitro* was comparable to that of sensory-evoked EPSCs *in vivo* (Fig. 4c). Despite substantial depression of peak amplitudes during high-frequency trains, the net charge transfer at the mossy fibre–granule cell synapse was affected remarkably little, with only about 30% depression observed at 500 Hz (Fig. 4d). The robustness of transmission at high frequencies was ensured in part by summation of the tails of EPSCs at high frequencies (Fig. 4e), most probably as a result of glutamate spillover from neighbouring synaptic contacts[16]. Consistent with this idea was the observation that the rise time of EPSCs and the proportion of slow-rising spillover EPSCs increased during both high-frequency stimulus-evoked EPSC trains *in vitro* and sensory-evoked EPSC bursts *in vivo* (Supplementary Fig. 5). These results demonstrate that a single mossy fibre bouton is capable of providing the rapid bursts of synaptic input to a granule cell observed after sensory stimulation.

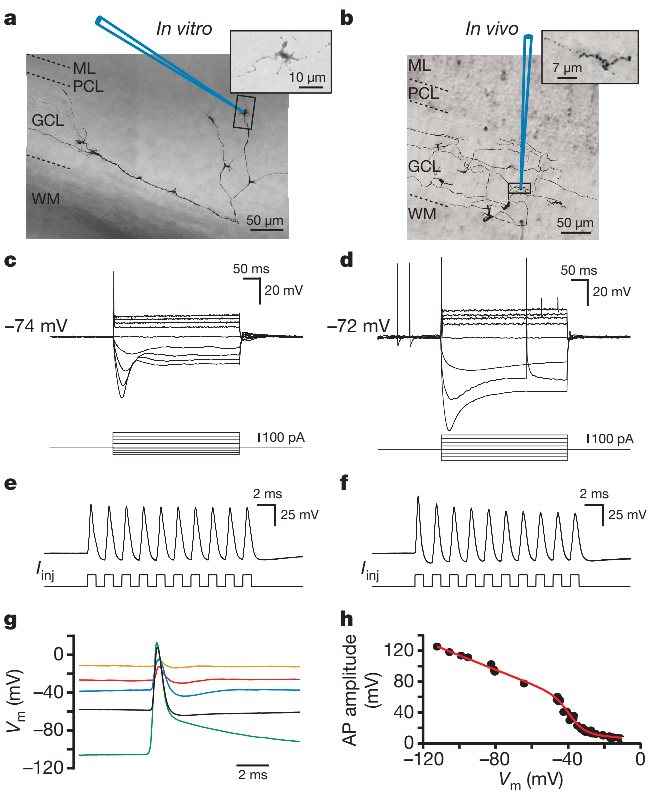


Figure 1 | Properties of mossy fibre boutons *in vitro* and *in vivo*. **a**, **b**, Whole-cell recordings were made from visually identified mossy fibre boutons in the granule cell layer of rat cerebellar slices by using infrared differential interference contrast microscopy (**a**) or blind-patch recording techniques *in vivo* (**b**). Morphological reconstructions of individual mossy fibres were obtained by biocytin labelling through the recording electrode. The schematic recording pipettes indicate the locations of the mossy fibre bouton recordings (*in vivo*: 549 μm from the pial surface). ML, molecular layer; PCL, Purkinje cell layer; GCL, granule cell layer; WM, white matter. Insets: fine structure of single 'mossy' ramifications. **c**, **d**, Mossy fibre boutons *in vitro* (**c**) and *in vivo* (**d**) show outward rectification and membrane potential sag in response to current injection. The occurrence of spontaneous action potentials *in vivo* is not affected by changes in membrane potential. **e**, **f**, Mossy fibre boutons can be driven to fire action potentials at extremely high frequencies by means of pulsed current injection (1 nA at 500 Hz); *in vitro* (**e**) $V_m = -58$ mV, *in vivo* (**f**) $V_m = -72$ mV, three consecutive traces. **g**, Representative spontaneous action potentials recorded at different membrane potentials *in vivo* (black trace at resting V_m). **h**, Relationship between spontaneous action potential amplitude and membrane potential (with the use of different levels of holding current) for a single mossy fibre bouton *in vivo*. The fit is a sigmoid function multiplied by a linear function to represent Na^+ channel availability and driving force, respectively. **a**, **c** and **e** were from the same recordings, as were **d**, **f**, **g** and **h**.

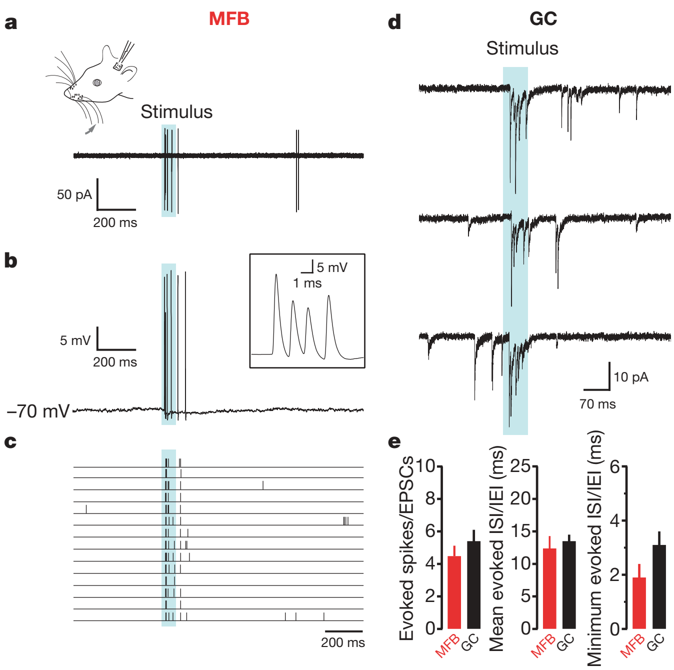


Figure 2 | Sensory-evoked spiking in mossy fibre boutons and EPSCs in granule cells *in vivo*. **a**, **b**, Cell-attached (**a**) and subsequent whole-cell (**b**) recording from a single mossy fibre bouton during stimulation with air puff (blue shaded area) of the upper lip area and whiskers. Inset: close-up of the high-frequency action potential burst evoked by sensory stimulation. Instantaneous spike frequencies were 658, 735 and 543 Hz, respectively. **c**, Raster plot of sensory-evoked action potentials (14 consecutive trials). **d**, EPSC bursts evoked by sensory stimulation recorded from a granule cell held at −70 mV (three consecutive trials) **e**, Number of presynaptic spikes and postsynaptic EPSCs and mean and minimum ISIs and IEIs evoked in mossy fibre boutons and granule cells by sensory stimulation ($n = 3$ and $n = 14$, respectively; $P > 0.05$). Even when correcting for synaptic failures (from Fig. 4), these values were not significantly different (see Methods).

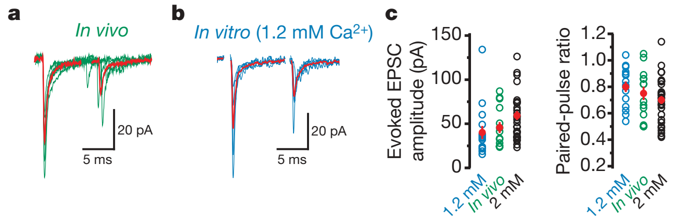


Figure 3 | Synaptic dynamics at the mossy fibre–granule cell synapse *in vivo* and *in vitro*. **a**, Sample EPSC traces evoked by whisker stimulation *in vivo*, aligned to the onset of the first EPSCs. Red traces display the peak-aligned average of the individual trials shown. **b**, Sample EPSC traces evoked by minimal stimulation *in vitro* in 1.2 mM $[Ca^{2+}]_e$. Red traces display the average of the individual trials shown; stimulus artefacts were removed for clarity. **c**, EPSC amplitudes and paired-pulse ratios of experiments *in vivo* (green, $n = 14$) and *in vitro* (1.2 mM $[Ca^{2+}]_e$, blue, $n = 16$; 2 mM $[Ca^{2+}]_e$, black, $n = 28$). The red symbols show means ± s.e.m. $V_{hold} = -70\,mV$ in **a** and **b**.

To test whether bursts of synaptic input provided by a single mossy fibre can trigger action potentials in granule cells, we stimulated single mossy fibre inputs (Supplementary Fig. 4) using patterns recorded from mossy fibre boutons *in vivo* (Fig. 5b), or using a regular stimulus pattern (five stimuli at 100 Hz) and recorded the resulting postsynaptic EPSPs and action potentials from postsynaptic granule cells *in vitro*. As the resting membrane potential of granule cells is variable *in vivo*[3], we measured the spiking output in response to the stimulus over a range of membrane potentials (Fig. 5a). Granule cells responded with bursts of spikes similar to that evoked *in vivo* with sensory stimulation[3], regardless of the extracellular Ca^{2+} concentration or the pattern of presynaptic burst activity (Fig. 5c), confirming that a burst of spikes can be triggered in granule cells by means of a single mossy fibre input.

We have made intracellular recordings from nerve terminals in the intact mammalian brain. Previous recordings *in vitro* have shown that boutons from hippocampal mossy fibres[6,17], basket cells[18] and the calyx of Held[19] show qualitatively similar excitabilities, in particular strong outward rectification. Our results further show that mossy fibre boutons *in vivo* are capable of extremely high maximal firing rates, and that such high firing rates are triggered by sensory stimulation. The sensory-evoked burst of spikes observed in mossy fibre boutons is consistent with rapid burst firing recorded from the somata of neurons that form the mossy fibres in the trigeminal nucleus[20], revealing that these high-frequency bursts are reliably transmitted from the soma to the axon terminals many millimetres distant. Our findings suggest that the release properties of mossy fibre boutons are matched to their high firing rates, and robust synaptic charge transfer is achievable at 500 Hz. The high fidelity of mossy fibre transmission is a consequence of several features of this synapse, including the multiple morphological contacts made by a mossy fibre with each granule cell[21], the rapid recovery of each release site from synaptic release[22], and the increasing contribution of glutamate spillover[16] at high frequencies.

The prevailing view of transmission at the relay from mossy fibre to granule cell is that activity in multiple mossy fibres is required to drive spike output[23–26]. Our findings provide direct support for a fundamentally different mode of operation of this relay, in which the sensory-evoked spiking observed in granule cells[3] can arise from a burst of spikes in a single mossy fibre, in an analogous manner to the 'detonator' function proposed for mossy fibres in the hippocampus[27,28]. Although integration of multiple mossy fibre inputs is likely to occur under some circumstances[23–26], our findings show that it is not a prerequisite for sensory activation of granule cells *in vivo*, consistent with Marr's proposal[25] that under conditions of reduced input from across a population of mossy fibres, granule cell output can be driven by a single presynaptic mossy fibre. Given that bursts of EPSCs are required to trigger spike output from granule cells *in vivo*[3], such an arrangement is ideally suited for the transfer of sensory information coded by a single sensory modality with the maximal signal-to-noise ratio. Such a single-fibre relay also provides the ultimate sensitivity to sensory input, yet it can reject spontaneous

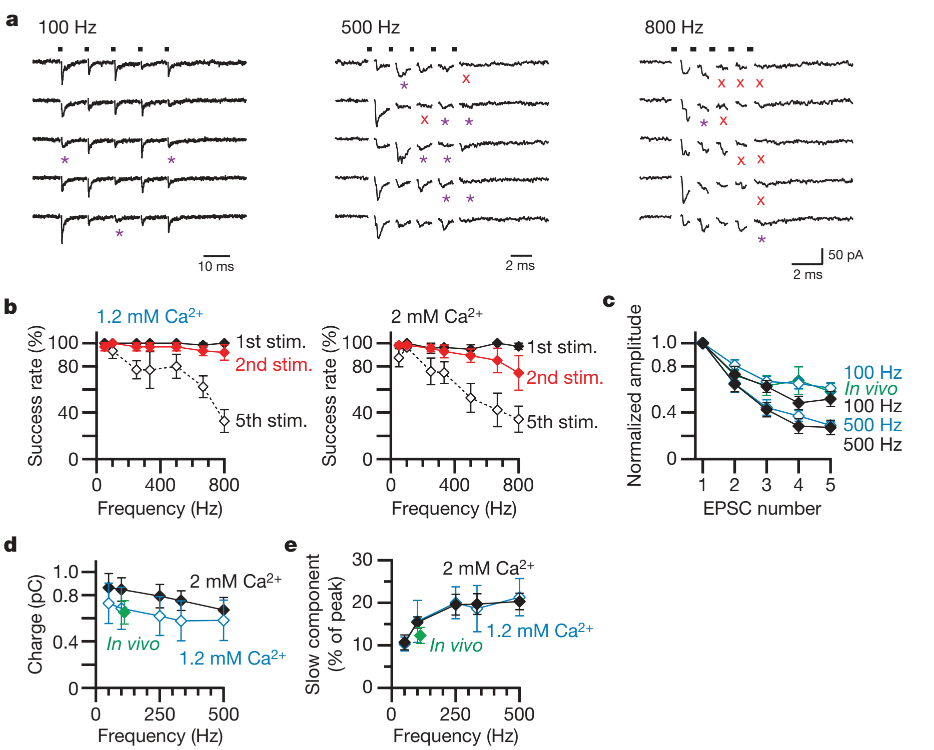


Figure 4 | High-fidelity synaptic transmission at the mossy fibre–granule cell synapse *in vivo* and *in vitro*. **a**, Sample EPSC traces evoked by trains of five stimuli at various frequencies *in vitro* in 1.2 mM $[Ca^{2+}]_e$. Five consecutive trials are shown. The dots at the top indicate the timing of the stimuli, the red crosses indicate failures of transmission and the purple asterisks denote slow-rising 'spillover' EPSCs. **b**, The probability of success (non-failure) for the first, second and fifth stimuli (stim.) at various frequencies (50–800 Hz). Data were obtained *in vitro* in 1.2 mM $[Ca^{2+}]_e$ (left, $n = 6$) and 2 mM $[Ca^{2+}]_e$ (right, $n = 7$–11). **c**, Synaptic depression of EPSCs during trains of five stimuli at 100 and 500 Hz *in vitro* and sensory-evoked EPSCs *in vivo*. Pooled data from 11 cells *in vitro* in 2 mM $[Ca^{2+}]_e$, (filled black symbols), six cells *in vitro* in 1.2 mM $[Ca^{2+}]_e$ (open blue symbols) and 14 cells *in vivo* (filled green symbols). **d**, Total EPSC charge transferred by five stimuli at various frequencies *in vitro* in 2 mM $[Ca^{2+}]_e$ ($n = 11$, filled black symbols) and in 1.2 mM $[Ca^{2+}]_e$ ($n = 6$, open blue symbols). Total charge of sensory-evoked EPSCs *in vivo* ($n = 14$, filled green symbol; horizontal error bars are smaller than the symbol) is plotted for the average frequency. **e**, The amplitudes of the slow-current component measured at 10 ms after the last (fifth) stimulus were normalized by the amplitude of the first EPSC. Pooled *in vitro* data in 2 mM $[Ca^{2+}]_e$ (filled black symbols, $n = 9$–11) and in 1.2 mM $[Ca^{2+}]_e$ (open blue symbols, $n = 6$) and *in vivo* data (filled green symbol, $n = 14$). Results are shown as means ± s.e.m.

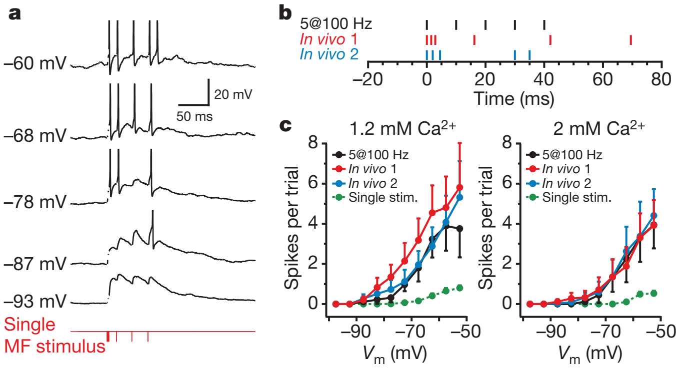


Figure 5 | Input from a single mossy fibre reliably drives granule cell firing. **a**, Stimulation of a single mossy fibre with a sensory-evoked stimulus pattern recorded *in vivo* (labelled '*In vivo* 1' in **b**) triggers bursts of action potentials in a granule cell *in vitro* (in 1.2 mM $[Ca^{2+}]_e$). Several sweeps at different membrane potentials (set by adjusting holding current) are shown (same cell as in Supplementary Fig. 4). Spikes are truncated and stimulus artefacts have been removed for clarity. **b**, Schematic representation of three different stimulation patterns used in **c** to activate single mossy fibre inputs (*in vivo* 1 and *in vivo* 2 were taken from different mossy fibre bouton recordings *in vivo*). **c**, Pooled data showing the number of action potentials triggered at different membrane potentials in 1.2 mM $[Ca^{2+}]_e$ (left) and 2 mM $[Ca^{2+}]_e$ (right).

synaptic input[3] and remains modifiable because the number and timing of granule cell output spikes can be regulated by inhibition[3,29]. The large synaptic divergence of single mossy fibre boutons[21], and the multiple collaterals shown by each mossy fibre[2], suggest that a sensory-evoked burst in a single mossy fibre may lead to the activation of a substantial number of neighbouring granule cells. Given that coincident activation of multiple granule cells is required for reliable activation of the downstream Purkinje cells[30], this arrangement thus solves the problem of generating synchrony in a population of granule cells with sparse afferent input.

METHODS SUMMARY

In vivo patch-clamp recordings were made from mossy fibre boutons and granule cells in folia Crus I and IIa of the cerebellar cortex of freely breathing 18–27-day-old Sprague–Dawley rats anaesthetized with urethane or with a ketamine/xylazine mixture as described previously[3]. Sensory responses were evoked by a brief air puff delivered to the ipsilateral whiskers or perioral surface[3]. Patch-clamp recordings from mossy fibre boutons and granule cells *in vitro* were made at 35–36 °C in cerebellar slices (200 μm thick) prepared with standard techniques[17]. Mossy fibre boutons were revealed by infrared differential interference contrast microscopy, with identification confirmed by biocytin staining after each experiment. For both *in vivo* and *in vitro* experiments, patch pipettes (6–9 MΩ for bouton recordings, and 5–8 MΩ for granule cell recordings) were filled with a potassium methanesulphonate-based internal solution. Data are given as means ± s.e.m.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 May; accepted 11 October 2007.

1. Ramón y Cajal, S. *La Textura del Sistema Nerviosa del Hombre y los Vertebrados* (Moya, Madrid, 1904).
2. Palay, S. L. & Chan-Palay, V. *Cerebellar Cortex—Cytology and Organization* (Springer, Berlin, 1974).
3. Chadderton, P., Margrie, T. W. & Häusser, M. Integration of quanta in cerebellar granule cells during sensory processing. *Nature* **428**, 856–860 (2004).
4. Holtzman, T., Rajapaksa, T., Mostofi, A. & Edgley, S. A. Different responses of rat cerebellar Purkinje cells and Golgi cells evoked by widespread convergent sensory inputs. *J. Physiol. (Lond.)* **574**, 491–507 (2006).
5. Simpson, J. I., Hulscher, H. C., Sabel-Goedknegt, E. & Ruigrok, T. J. Between in and out: linking morphology and physiology of cerebellar cortical interneurons. *Prog. Brain Res.* **148**, 329–340 (2005).
6. Engel, D. & Jonas, P. Presynaptic action potential amplification by voltage-gated Na^+ channels in hippocampal mossy fiber boutons. *Neuron* **45**, 405–417 (2005).
7. Hodgkin, A. L. & Huxley, A. F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol. (Lond.)* **117**, 500–544 (1952).
8. Eccles, J. C., Faber, D. S., Murphy, J. T., Sabah, N. H. & Taborikova, H. Afferent volleys in limb nerves influencing impulse discharges in cerebellar cortex. I. In mossy fibers and granule cells. *Exp. Brain Res.* **13**, 15–35 (1971).
9. van Kan, P. L., Gibson, A. R. & Houk, J. C. Movement-related inputs to intermediate cerebellum of the monkey. *J. Neurophysiol.* **69**, 74–94 (1993).
10. Garwicz, M., Jörntell, H. & Ekerot, C. F. Cutaneous receptive fields and topography of mossy fibres and climbing fibres projecting to cat cerebellar C3 zone. *J. Physiol. (Lond.)* **512**, 277–293 (1998).
11. Lisberger, S. G. & Fuchs, A. F. Role of primate flocculus during rapid behavioral modification of vestibuloocular reflex. II. Mossy fiber firing patterns during horizontal head rotation and eye movement. *J. Neurophysiol.* **41**, 764–777 (1978).
12. Morissette, J. & Bower, J. M. Contribution of somatosensory cortex to responses in the rat cerebellar granule cell layer following peripheral tactile stimulation. *Exp. Brain Res.* **109**, 240–250 (1996).
13. Silver, R. A., Cull-Candy, S. G. & Takahashi, T. Non-NMDA glutamate receptor occupancy and open probability at a rat cerebellar synapse with single and multiple release sites. *J. Physiol. (Lond.)* **494**, 231–250 (1996).
14. Silver, I. A. & Erecinska, M. Intracellular and extracellular changes of $[Ca^{2+}]$ in hypoxia and ischemia in rat brain *in vivo*. *J. Gen. Physiol.* **95**, 837–866 (1990).
15. Kristian, T. & Siesjo, B. K. Calcium in ischemic cell death. *Stroke* **29**, 705–718 (1998).
16. DiGregorio, D. A., Nusser, Z. & Silver, R. A. Spillover of glutamate onto synaptic AMPA receptors enhances fast transmission at a cerebellar synapse. *Neuron* **35**, 521–533 (2002).
17. Geiger, J. R. P. & Jonas, P. Dynamic control of presynaptic Ca^{2+} inflow by fast-inactivating K^+ channels in hippocampal mossy fiber boutons. *Neuron* **28**, 927–939 (2000).
18. Southan, A. P., Morris, N. P., Stephens, G. J. & Robertson, B. Hyperpolarization-activated currents in presynaptic terminals of mouse cerebellar basket cells. *J. Physiol. (Lond.)* **526**, 91–97 (2000).
19. Cuttle, M. F., Rusznák, Z., Wong, A. Y., Owens, S. & Forsythe, I. D. Modulation of a presynaptic hyperpolarization-activated cationic current (I_h) at an excitatory synaptic terminal in the rat auditory brainstem. *J. Physiol. (Lond.)* **534**, 733–744 (2001).
20. Jones, L. M., Lee, S., Trageser, J. C., Simons, D. J. & Keller, A. Precise temporal responses in whisker trigeminal neurons. *J. Neurophysiol.* **92**, 665–668 (2004).
21. Jakab, R. L. & Hamori, J. Quantitative morphology and synaptology of cerebellar glomeruli in the rat. *Anat. Embryol. (Berl.)* **179**, 81–88 (1988).
22. Saviane, C. & Silver, R. A. Fast vesicle reloading and a large pool sustain high bandwidth transmission at a central synapse. *Nature* **439**, 983–987 (2006).
23. Albus, J. S. A theory of cerebellar function. *Math. Biosci.* **10**, 25–61 (1971).
24. D'Angelo, E., De Filippi, G., Rossi, P. & Taglietti, V. Synaptic excitation of individual rat cerebellar granule cells *in situ*: evidence for the role of NMDA receptors. *J. Physiol. (Lond.)* **484**, 397–413 (1995).
25. Marr, D. A theory of cerebellar cortex. *J. Physiol. (Lond.)* **202**, 437–470 (1969).
26. Jörntell, H. & Ekerot, C. F. Properties of somatosensory synaptic integration in cerebellar granule cells *in vivo*. *J. Neurosci.* **26**, 11786–11797 (2006).
27. McNaughton, B. L. & Morris, R. G. M. Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends Neurosci.* **10**, 408–415 (1987).
28. Henze, D. A., Wittner, L. & Buzsáki, G. Single granule cells reliably discharge targets in the hippocampal CA3 network *in vivo*. *Nature Neurosci.* **5**, 790–795 (2002).
29. Maex, R. & De Schutter, E. Synchronization of Golgi and granule cell firing in a detailed network model of the cerebellar granule cell layer. *J. Neurophysiol.* **80**, 2521–2537 (1998).
30. Barbour, B. Synaptic currents evoked in Purkinje cells by stimulating individual granule cells. *Neuron* **11**, 759–769 (1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Geiger for guidance during the slice experiments; B. Clark, J. Davie, M. Farrant and A. Roth for comments on the manuscript; K. Kitamura, S. Komai and M. Rizzi for help with preliminary experiments; and L. Ramakrishnan and H. Cuntz for help with histology. This work was supported by grants from the European Union, Wellcome Trust and Gatsby Foundation (M.H.), and by a Wellcome Prize Studentship (E.A.R.), a Human Frontier Science Program Long-Term Fellowship (T.I.), a Wellcome Trust Advanced Training Fellowship (I.D.) and a University College London Graduate School Research Scholarship (P.C.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.H. (m.hausser@ucl.ac.uk).

METHODS

The care and experimental manipulation of animals were conducted in accordance with institutional and national guidelines.

General. Current-clamp and voltage-clamp recordings were made with Multiclamp 700A and 700B amplifiers (Molecular Devices). The internal solution contained (in mM): potassium methanesulphonate 133, KCl 7, HEPES 10, MgATP 2, Na_2ATP 2, Na_2GTP 0.5, EGTA 0.05 (pH 7.2). Biocytin (0.5%) was added for subsequent morphological reconstruction. Data were filtered at 3–10 kHz and acquired at 50 kHz using Axograph software (http://axographx.com/) in conjunction with an ITC-18 interface (Instrutech). Seal resistance was always more than 3 GΩ in the cell-attached configuration. Resting membrane potentials (V_m) were measured immediately after formation of the whole-cell configuration ('break-in'). The series resistance for *in vitro* and *in vivo* mossy fibre terminal recordings was 35 ± 10 MΩ ($n = 8$) and 72 ± 20 MΩ ($n = 4$), determined from the passive current response to a voltage step at the end of the recording (see Supplementary Fig. 2). For recordings from granule cells, the series resistance was 32 ± 3 MΩ ($n = 44$) and 33 ± 2 MΩ ($n = 14$) for *in vitro* and *in vivo* recordings, respectively. Capacitive currents recorded from mossy fibre terminals in response to voltage steps were described by a double-exponential function (Supplementary Fig. 2), with the dominant fast component presumably corresponding to the charging of the terminal, and the slower component representing the axon[17]. The fast component was used to estimate the capacitance of the terminal.

***In vivo* recordings.** Sprague–Dawley rats (18–27 days old) were anaesthetized with urethane (1.2 g kg^{-1}) or with a ketamine (50 mg kg^{-1})/xylazine (5 mg kg^{-1}) mixture as described previously[3,31]. Recordings were made more than 400 μm from the pial surface in the granule cell layer. Sensory stimulation was performed with an airpuff (30–70 ms, 60 lb in^{-2} (approx. 410 kPa)) timed by a Picospritzer (General Valve) and aimed at the ipsilateral whiskers or perioral surface with a glass tube mounted on a micromanipulator.

***In vitro* recordings.** Recordings from mossy fibre boutons in slices were made with standard techniques[32]. The external solution contained (in mM): NaCl 125, $NaHCO_3$ 25, glucose 25, KCl 2.5, NaH_2PO_4 1.25, $CaCl_2$ 2, $MgCl_2$ 1 (pH 7.3 when bubbled with 95% O_2 and 5% CO_2). In some experiments, $CaCl_2$ concentration was decreased to 1.2 mM as described, because previous reports[14,15,33] had indicated that $[Ca^{2+}]_e$ in the cerebrospinal fluid may be somewhat lower than that widely used for *in vitro* slice experiments. The recording chamber was continuously perfused with external solutions and maintained at physiological temperature (35–36 °C). EPSCs were evoked by extracellular stimulation (100 μs, typically 5–15 V) at 0.5 Hz, unless otherwise stated, using a monopolar electrode or a bipolar electrode made from a theta capillary filled with external solution and placed in the granule cell layer, about 50 μm from the recording site. EPSCs from single mossy fibre inputs were identified by their all-or-none appearance when the stimulation strength was gradually increased[13] (see Supplementary Fig. 4).

Analysis. Data analysis was performed with Axograph and Igor Pro (Wavemetrics). Input resistance and rectification ratio were calculated from the voltage response to symmetrical hyperpolarizing or depolarizing current steps (400 ms; 10–40 pA) from rest, measured at the end of the step. Sag ratio was calculated as the peak divided by the steady-state input resistance from hyperpolarizing voltage deflections reaching −100 mV. Spike broadening *in vivo* was calculated by normalizing the half-width of the spike following the shortest ISI in the sensory-evoked burst by the half-width of the first spike. Spike broadening *in vitro* was measured with current-injection-evoked spike trains (five at 100 Hz), and the half-width of the fifth spike was normalized by the half-width of the first spike. EPSC traces were digitally low-pass filtered (3 kHz). The latency of sensory-evoked presynaptic action potentials and postsynaptic EPSCs were corrected to account for the temporal delay of the air puff associated with the application system. For stimulus-evoked EPSCs, synaptic responses were categorized as successes when their amplitudes were larger than a threshold (typically about 7 pA), defined as 4.5-fold the standard deviation of the baseline noise. Synaptic charge measurements were made by using averaged traces from which artefacts were removed by linear interpolation between points immediately before and after the artefacts. To identify 'spillover' EPSCs[16] *in vitro* and *in vivo*, events were defined as slow-rising spillover EPSCs if the 20–80% rise time was longer than 1.6-fold the median of the rise-time distribution obtained from the same cell. This cut-off value roughly corresponded to the mean plus four times the standard deviation because the coefficient of variation of the rise-time distribution was typically 0.15 in cells that did not show obvious slow-rising events. These criteria corresponded well to events categorized by eye. Statistical comparisons were made with Student's two-sided paired *t*-test unless otherwise indicated.

The effect of synaptic failures on *in vivo* sensory-evoked EPSCs was evaluated by numerical simulation with the sensory-evoked spiking pattern of *in vivo* mossy fibre boutons (Fig. 2) and the average failure rate estimated from mossy fibre stimulation experiments (Fig. 4b; 1.2 mM $[Ca^{2+}]_e$ *in vitro*). For each presynaptic action potential, a failure rate was defined by using a value corresponding to the nearest ISI shown in Fig. 4b. Values from 1,000 random trials were averaged for each cell. As described in the results and in Fig. 2, the number, minimum interval and mean interval of sensory-evoked presynaptic spikes in single boutons were 4.2 ± 0.6, 1.9 ± 0.5 ms and 12.4 ± 1.7 ms, respectively ($n = 3$). If there are no failures and there is perfectly reliable transmission across the synapse, the corresponding values for postsynaptic sensory-evoked EPSCs should be identical (assuming that only one bouton is involved). Taking into account the failure rate, the number, minimum interval and mean interval of EPSCs were predicted to be 3.9 ± 0.5, 1.9 ± 0.5 ms and 13.0 ± 1.7 ms, respectively ($n = 3$). These values were similar to the number (5.4 ± 0.7; $P = 0.35$), the minimum interval (3.1 ± 0.5 ms; $P = 0.27$) and the mean interval (13.5 ± 1.0 ms; $P = 0.82$) of actual sensory-evoked EPSCs ($n = 14$, unpaired *t*-test). This result is consistent with our conclusion that the sensory-evoked EPSCs arise from activity in a single mossy fibre.

31. Margrie, T. W., Brecht, M. & Sakmann, B. *In vivo*, low-resistance, whole-cell recordings from neurons in the anaesthetized and awake mammalian brain. *Pflügers Arch.* **444**, 491–498 (2002).
32. Bischofberger, J., Engel, D., Li, L., Geiger, J. R. & Jonas, P. Patch-clamp recording from mossy fiber terminals in hippocampal slices. *Nature Protocols* **1**, 2075–2081 (2006).
33. Leusen, I. Regulation of cerebrospinal fluid composition with reference to breathing. *Physiol. Rev.* **52**, 1–56 (1972).

Vol 450 | 20/27 December 2007 | doi:10.1038/nature06395

LETTERS

Stochastic gene expression out-of-steady-state in the cyanobacterial circadian clock

Jeffrey R. Chabot[1]†, Juan M. Pedraza[1], Prashant Luitel[1] & Alexander van Oudenaarden[1]

Recent advances in measuring gene expression at the single-cell level have highlighted the stochastic nature of messenger RNA and protein synthesis[1–3]. Stochastic gene expression creates a source of variability in the abundance of cellular components, even among isogenic cells exposed to an identical environment. Recent integrated experimental and modelling studies[4–13] have shed light on the molecular sources of this variability. However, many of these studies focus on systems that have reached a steady state and therefore do not address a large class of dynamic phenomena including oscillatory gene expression. Here we develop a general protocol for analysing and predicting stochastic gene expression in systems that never reach steady states. We use this framework to analyse experimentally stochastic expression of genes driven by the *Synechococcus elongatus* circadian clock. We find that, although the average expression at two points in the circadian cycle separated by 12 hours is identical, the variability at these two time points can be different. We show that this is a general feature of out-of-steady-state systems. We demonstrate how intrinsic noise sources, owing to random births and deaths of mRNAs and proteins, or extrinsic noise sources, which introduce fluctuations in rate constants, affect the cell-to-cell variability. To distinguish experimentally between these sources, we measured how the correlation between expression fluctuations of two identical genes is modulated during the circadian cycle. This quantitative framework is generally applicable to any out-of-steady-state system and will be necessary for understanding the fidelity of dynamic cellular systems.

So far, the expression reporter used in the cyanobacterium *S. elongatus* PCC7942 has been bacterial luciferase[14,15]. This technique allows population-level measurements but limits studies of expression fluctuations between individuals, because of the very weak bioluminescence signal emitted from single cells[16]. We therefore developed a single-cell fluorescent reporter assay that provides a much stronger signal facilitating quick and accurate measurements of gene expression levels in large numbers of individual cells (Supplementary Fig. 1). We found that an SsrA-tagged[17] yellow-shifted variant of green fluorescent protein, YFP–SsrA(LVA), in which the last three amino acids in the SsrA tag are leucine, valine and alanine, was able to report faithfully the periodic activity of the *S. elongatus kaiBC* promoter (P_{kaiBC}) (Fig. 1, Supplementary Movie 1).

We used this fluorescence assay to measure YFP–SsrA(LVA) expression in individual *Synechococcus* cells using flow cytometry. Using two known neutral loci in the *Synechococcus* genome, defined as NS I and NS II (ref. 18), to insert the P_{kaiBC}–*yfp*–*ssrA(LVA)* construct, we found that the mean YFP–SsrA(LVA) expression (averaged over 10^5 individual cells) exhibits a clear periodic expression pattern with levels that are very similar between the two neutral sites (Fig. 2a). The phase, however, differs from previously reported mRNA oscillations, peaking approximately 4 to 6 h later than the mRNA maximum[14]. From the observed half-life of YFP–SsrA(LVA), we expect a phase delay of 4.3 h, consistent with the observed phase shift (Supplementary Information). Interestingly, the key circadian clock protein KaiC is similarly long-lived, with a half-life of about 10 h (ref. 19). This indicates that this endogenous protein will also display a similar phase shift (up to a maximum of 6 h) with respect to the phase of the corresponding *kaiC* mRNA concentration. Using the measured lifetime of YFP–SsrA(LVA) and an estimate for the half-life of the corresponding mRNA of about 15 min (ref. 19), the mRNA creation rate $k_R(t)$ can be inferred from the YFP–SsrA(LVA)

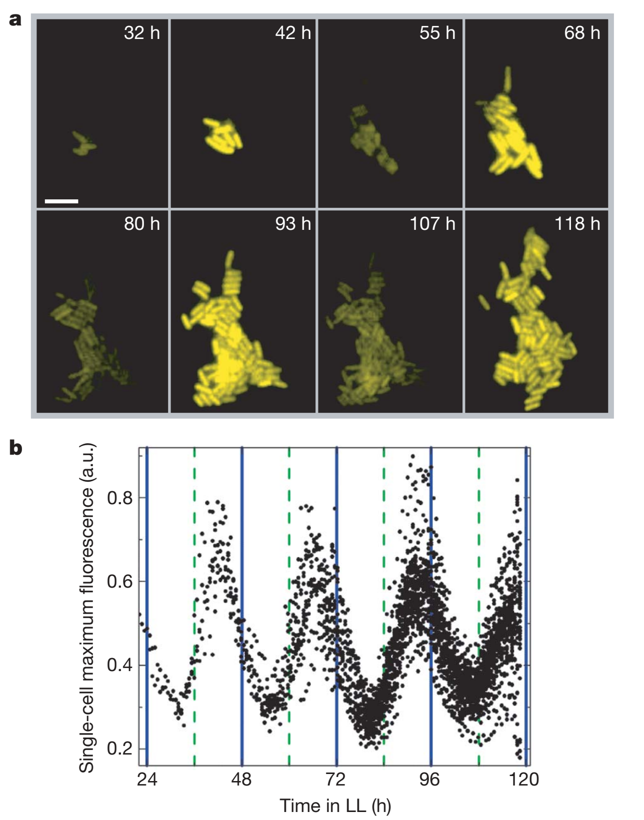


Figure 1 | Monitoring circadian oscillations in single *S. elongatus* PCC7942 cells using fluorescence microscopy. a, Montage of fluorescence microscopy images demonstrating circadian oscillations in single cells. Time is reported in hours (h). The scale bar represents 4 μm. **b,** Analysis of maximum pixel fluorescence within individual cells (measured in arbitrary units, a.u.) as a function of time. LL, constant light.

[1]Department of Physics and George Harrison Spectroscopy Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. †Present address: Systems Biology Group, Pfizer Research Technology Center, Cambridge, Massachusetts 02139, USA.

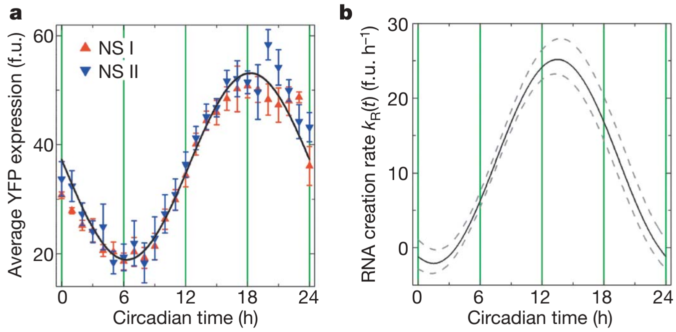


Figure 2 | Circadian oscillations in single cells. **a**, Average YFP–SsrA(LVA) expression as a function of circadian time for strain JRCS32 (red triangles) and JRCS35 (blue triangles). Fluorescence was obtained by averaging the fluorescence of at least 10^4 single cells. Cells were synchronized by two 12:12 LD cycles before the start of the experiment. The black solid line is a fit to both data sets with a cosine function with a period of 24 h. f.u., fluorescence units. Error bars, 1 s.e.m. **b**, Inferred RNA creation rate $k_R(t)$ as a function of circadian time. Dashed lines indicate the error bars given the experimental uncertainty of the YFP–SsrA(LVA) half-life (5.6 ± 1.0 h).

dynamics (Fig. 2b, Supplementary Information). We find that the *yfp–ssrA(LVA)* mRNA concentration peaks slightly after a circadian time of 12 h and vanishes 6 h before and after this peak consistent with northern blot analyses[14].

This single-cell assay allows us to probe expression variability quantitatively from cell to cell, and to explore the effect of stochastic gene expression in this oscillating system. The expression variability was quantified by the coefficient of variation, which is the standard deviation of the expression distribution normalized by the average expression. Figure 3a displays the squared coefficient of variation as a function of the inverse averaged expression. We chose this representation because in steady-state one would expect a linear relationship between these variables[8,20]. However, we find that during the circadian cycle the experimental data trace out a loop (Fig. 3a). In other words, for a particular average expression, two values of coefficient of variation are possible: a low variability early in the circadian cycle and a high variability later in the cycle. This property cannot be understood by previous theoretical work predicting the stochastic properties of gene expression assuming that protein and mRNA creation and destruction reached a steady state[8,20–22]. This assumption is violated in this circadian clock because the oscillation period is comparable to the protein half-life and therefore the system will never reach even a steady-state in which protein production is balanced with protein destruction. We therefore developed a new theoretical framework that allows us to calculate the cell-to-cell variability in systems that are not in steady-state. Using this model, we indeed find that the loops can be reproduced qualitatively using either an intrinsic noise model (Fig. 3b) or an extrinsic noise model (Fig. 3c). Remarkably, the chirality of the loop depends on how rapidly the extrinsic noise fluctuates.

To characterize the relative contributions of intrinsic and extrinsic sources experimentally, we devised a strategy to measure the correlation between expression fluctuations at NS I and those at NS II. A powerful strategy would be to use two different coloured fluorescent proteins at the two neutral sites and compare the expression values in single cells[5,7]. However, because YFP is the only reliable reporter in *Synechococcus*, we decided to take an alternative approach. We constructed a strain that contains P_{kaiBC}–*yfp–SsrA(LVA)* in both NS I and NS II. As expected, the average YFP expression doubles (Fig. 4a, black symbols). However the variability observed on doubling of the gene dose displays a more complicated behaviour (Fig. 4b).

The total variance observed for one copy is partly caused by a noise source that introduces correlated expression fluctuations between NS I and NS II, and partly caused by another noise source that introduces uncorrelated fluctuations. We will designate the correlated source as global and the uncorrelated source as local. These two sources contribute independently to the total variance, and therefore the total variance for a single copy is given by $\sigma_1^2(t)=\sigma_{\mathrm{local}}^2(t)+\sigma_{\mathrm{global}}^2(t)$. In the case of two copies, the local component of the variance will double and the global component will quadruple[23,24]: $\sigma_2^2(t)=2\sigma_{\mathrm{local}}^2(t)+4\sigma_{\mathrm{global}}^2(t)$. Therefore, by using these two equations, the local and global component of the variance can be explicitly found as a function of time (t) (Fig. 4c, d). We find that $\sigma_{\mathrm{local}}^2(t)$ and $\sigma_{\mathrm{global}}^2(t)$ have comparable magnitude but peak at different times of the day. In the next step we use the dynamics of the local and global component of the variance to infer what the

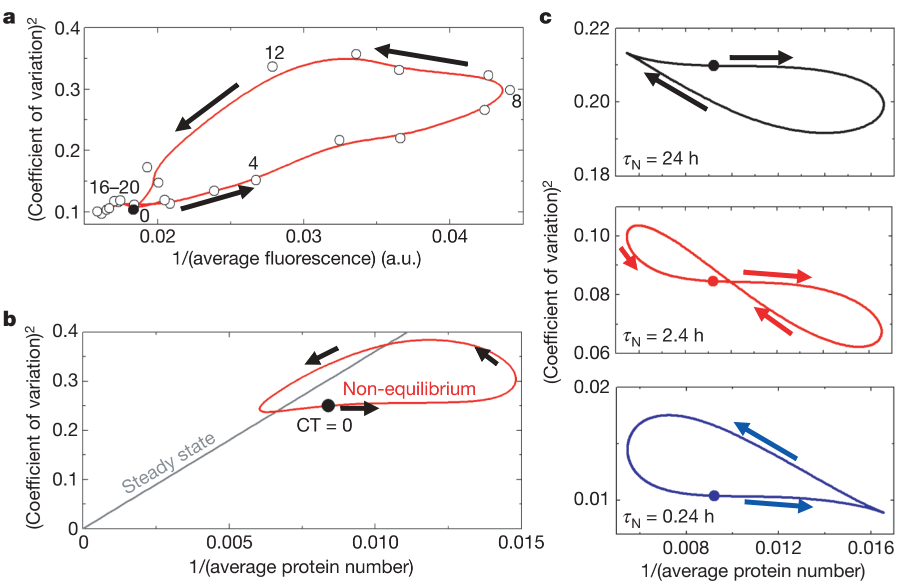


Figure 3 | A dynamic analysis of stochastic gene expression reveals noise loops. **a**, The squared coefficient of variation as a function of the inverse average fluorescence for strain JRCS32 (open circles). The filled circle denotes the start of the circadian cycle (circadian time = 0) and the numbers denote the circadian time. The red line serves as a guide to the eye. **b**, The squared coefficient of variation as a function of the inverse average protein number using an intrinsic noise model (red solid line). The linear relationship expected in steady state is shown in grey. **c**, The squared coefficient of variation as a function of the inverse average protein number using an extrinsic noise model (solid lines) for different noise decorrelation times ($\tau_N=\frac{\ln 2}{\gamma_N}$).

variability introduced at the transcriptional level had to be to explain the magnitude and dynamics of $\sigma^2_{\mathrm{local}}(t)$ and $\sigma^2_{\mathrm{global}}(t)$.

To determine the magnitude of the fluctuations introduced at the transcriptional level, we consider the following model. The dynamics of the *yfp–SsrA(LVA)* mRNA concentration (x) and YFP–SsrA(LVA) protein concentration (y) can be described generally by the following dynamical system of equations:

$$\frac{\mathrm{d}x}{\mathrm{d}t}=k_{\mathrm{R}}(t)-\gamma_{\mathrm{R}}x+q(t)\xi(t)$$

$$\frac{\mathrm{d}y}{\mathrm{d}t}=k_{\mathrm{P}}x-\gamma_{\mathrm{P}}y$$

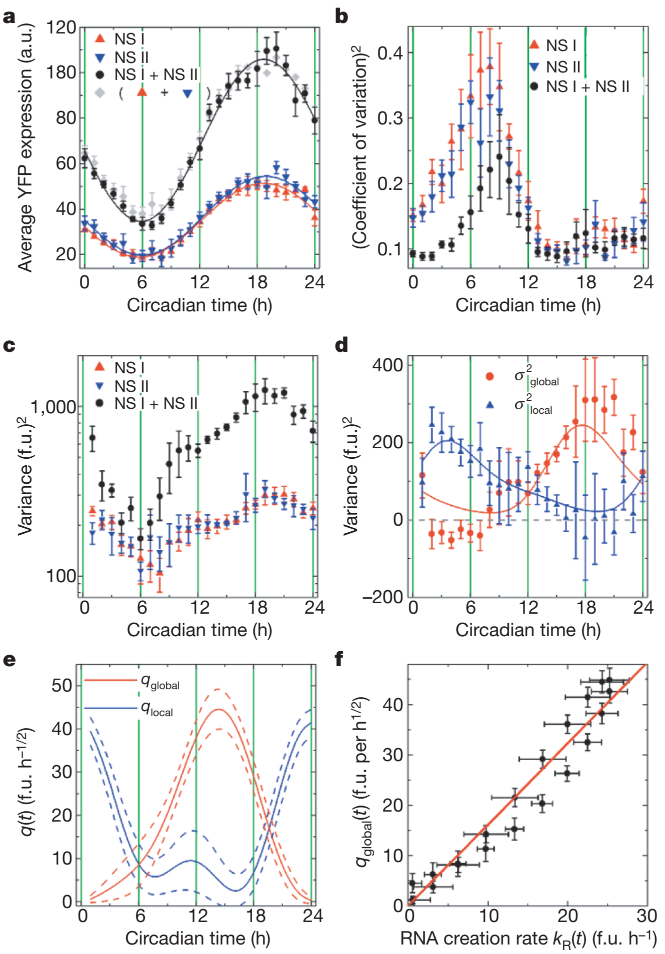


Figure 4 | Cell to cell variability in single- and double-copy-number constructs. a, Average YFP–SsrA(LVA) expression as a function of circadian time for strain JRCS32 (red triangles), JRCS35 (blue triangles) and JRCS70 (black circles). Fluorescence was obtained by averaging the fluorescence of at least 10^4 single cells. Cells were synchronized by two 12:12 LD cycles before the start of the experiment. Solid lines denote fits to a cosine function with a period of 24 h. The grey diamonds note the addition of the two fluorescence values of strain JRCS32 and JRCS35. Error bars denoted 1 s.e.m. obtained from day-to-day reproductions. **b,** The squared coefficient of variation as a function of circadian time. This was computed from at least 10^4 single cells. Error bars represent standard errors computed from day-to-day reproductions. **c,** The variance of the expression distribution as a function of circadian time using the same strains as for **a**. The variances were obtained by averaging the fluorescence of at least 10^4 single cells. Error bars (1 s.e.m) indicate day-to-day variation determined by repeated reproduction. **d,** Component of variance that reflects local (blue triangles) and global (red circles) fluctuations as a function of circadian time. Error bars (1 s.e.m) were obtained by error propagation. **e,** Computed functions $q_{\mathrm{local}}(t)$ (blue) and $q_{\mathrm{global}}(t)$ (red) as a function of circadian time. Dashed lines indicate error bars on the basis of error propagation. **f,** Correlation between $q_{\mathrm{global}}(t)$ and the RNA transcription rate $k_{\mathrm{R}}(t)$. The solid red line represents a linear fit.

where γ_{R} and γ_{P} are the rate constants for RNA degradation and protein degradation, respectively. The estimated mRNA half-life of 15 min (ref. 19) gives $\gamma_{\mathrm{R}}\sim 2.8\,\mathrm{h}^{-1}$. Using the experimentally determined YFP–SsrA(LVA) protein degradation rate (Supplementary Fig. 1) we find $\gamma_{\mathrm{P}}\sim 0.12\,\mathrm{h}^{-1}$. The rate constant for translation, k_{P}, is unknown. The function $k_{\mathrm{R}}(t)$ describes the periodic transcription rate. The absolute transcription rates are unknown; however, we obtained a good estimate of the relative dynamics (Fig. 2b). The noise term $\xi(t)$ reflects gaussian noise with zero mean, $\langle\xi(t)\rangle=0$, and exponentially decaying correlation, $\langle\xi(t)\xi(\tau)\rangle=\frac{\gamma_{\mathrm{N}}}{2}\mathrm{e}^{-\gamma_{\mathrm{N}}|t-\tau|}$, with decay rate γ_{N}. This type of noise is consistent with recent experimental data[25] that show that γ_{N} is of the order of the cell-doubling rate. We assume that all the stochastic fluctuations are introduced at the level of mRNA production/decay, and that protein production/decay follows deterministic dynamics. Using this model we can determine $q(t)$ from the experimentally determined variances (Fig. 4c, d). The function $q(t)$ is a periodic function that describes the magnitude of the stochastic fluctuations introduced at the transcriptional level. Similarly to the separation of the total variance in terms of $\sigma^2_{\mathrm{local}}(t)$ and $\sigma^2_{\mathrm{global}}(t)$, we can separate $q(t)\xi(t)$ into a local and a global component: $q_{\mathrm{local}}(t)\xi_{\mathrm{l}}(t)$ and $q_{\mathrm{global}}(t)\xi_{\mathrm{g}}(t)$, which can be determined from the experimentally obtained $\sigma^2_{\mathrm{local}}(t)$ and $\sigma^2_{\mathrm{global}}(t)$ (Fig. 4e).

We find that $q_{\mathrm{global}}(t)$ reaches a maximum around a circadian time of 14 h. These dynamics closely resemble the dynamics of $k_{\mathrm{R}}(t)$ (Fig. 2b). Figure 4f demonstrates explicitly the strong correlation between $k_{\mathrm{R}}(t)$ and $q_{\mathrm{global}}(t)$. This correlation is consistent with global fluctuations introduced at the level of RNA creation (Supplementary Information): $q_{\mathrm{global}}(t)=\sqrt{2/\gamma_{\mathrm{N}}}\eta_{\mathrm{G}}k_{\mathrm{R}}(t)$. We use the slope of the $k_{\mathrm{R}}(t)-q_{\mathrm{global}}(t)$ curve (Fig. 4f) to estimate the magnitude η_{G} of these global fluctuations: $\eta_{\mathrm{G}}\sim 0.4$. Here we assume that the noise decay rate γ_{N} is set by the rate constant of protein destruction γ_{P}. This indicates that at any particular circadian time the global transcription rate of the *kaiBC* promoter varies about 40% from cell to cell.

The function $q_{\mathrm{local}}(t)$ reaches a maximum around a circadian time of 0 h. A potential candidate for the local noise source could be intrinsic fluctuations caused by random births and deaths of mRNA molecules. However if $q_{\mathrm{local}}(t)$ would be dominated by intrinsic noise, one would expect $q_{\mathrm{local}}(t)\propto\sqrt{k_{\mathrm{R}}(t)}$, and therefore $q_{\mathrm{local}}(t)$ should peak around a circadian time of 14 h (see ref. 26), which is not observed (Fig. 4e). This rules out not only intrinsic noise but also a wide class of mechanisms, because any noise that is a monotonically increasing function of transcription rate is inconsistent with this observation. One potential candidate is the difference in local cellular environments between the two neutral sites. This hypothesis is consistent with the 'oscilloid' model[27] in which the condensation or supercoiling state of the cyanobacterial chromosome changes with the circadian rhythm. Indeed, recent experimental evidence has demonstrated a periodic cycle of compaction and relaxation of the cyanobacterial chromosome[28].

How fluctuations in clock components affect the fidelity of an oscillator has been modelled in detail[29,30]. Here we explore the stochastic expression of a gene driven by the clock. We show that in oscillatory systems a phase difference arises between signal and noise. This results in noise loops that cannot be explained by current steady-state models. This quantitative framework expands the current stochastic analysis toolbox to include a large class of non-steady-state phenomena such as oscillatory and transient dynamics.

METHODS SUMMARY

Growth media and cell culture. All strains were grown at 30 °C in BG11 media, supplemented with appropriate antibiotics (40 μg ml^{-1} spectinomycin for NS I inserts; 50 μg ml^{-1} kanamycin for NS II inserts). 'Light' environments were maintained at ~2,400 lx (lux) from soft-fluorescent sources (Sylvania); cells in the 'dark' environments experienced less than 50 lx. To synchronize samples, cells were exposed to at least two full 12 h dark:12 h light (12:12 LD) cycles before beginning data collection.

Out-of-steady-state model for stochastic gene expression. It can be shown (Supplementary Information) that the variance in the mRNA concentration $\sigma_x^2(t)$ is given by the following convolution: $\sigma_x^2(t) = \langle (q(t)\xi(t) \otimes f(t))^2 \rangle$, where $f(t) = \exp[-\gamma_R t]$. Similarly, the variance in the protein concentration $\sigma_y^2(t)$ can be calculated: $\sigma_y^2(t) = k_P^2 \langle (q(t)\xi(t) \otimes g(t))^2 \rangle$, where $g(t) = \frac{\exp[-\gamma_P t]}{\gamma_R - \gamma_P} + \frac{\exp[-\gamma_R t]}{\gamma_P - \gamma_R}$. This procedure can be applied to $\sigma_{global}^2(t)$ and $\sigma_{local}^2(t)$ (Fig. 4d) to determine $q_{global}(t)$ and $q_{local}(t)$ (Fig. 4e):

$$\sigma_{global}^2(t) = \frac{1}{2}\left[\sigma_2^2(t) - 2\sigma_1^2(t)\right] = k_P^2 \left\langle \left(q_{global}(t)\xi(t) \otimes g(t)\right)^2 \right\rangle$$

$$\sigma_{local}^2(t) = 2\sigma_1^2(t) - \frac{1}{2}\sigma_2^2(t) = k_2 P^2 \langle (q_{local}(t)\xi(t) \otimes g(t))^2 \rangle$$

Received 13 August; accepted 16 October 2007.

1. Kaern, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: from theories to phenotypes. *Nature Rev. Genet.* **6**, 451–464 (2005).
2. Kaufmann, B. B. & van Oudenaarden, A. Stochastic gene expression: from single molecules to the proteome. *Curr. Opin. Genet. Dev.* **17**, 107–112 (2007).
3. Maheshri, N. & O'Shea, E. K. Living with noisy genes: how cells function reliably with inherent variability in gene expression. *Annu. Rev. Biophys. Biomol. Struct.* **36**, 413–434 (2007).
4. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nature Genet.* **31**, 69–73 (2002).
5. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
6. Blake, W. J., Kaern, M., Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression. *Nature* **422**, 633–637 (2003).
7. Raser, J. M. & O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811–1814 (2004).
8. Paulsson, J. Summing up the noise in gene networks. *Nature* **427**, 415–418 (2004).
9. Cai, L., Friedman, N. & Xie, X. S. Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**, 358–362 (2006).
10. Newman, J. R. *et al.* Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* **441**, 840–846 (2006).
11. Bar-Even, A. *et al.* Noise in protein expression scales with natural protein abundance. *Nature Genet.* **38**, 636–643 (2006).
12. Sigal, A. *et al.* Variability and memory of protein levels in human cells. *Nature* **444**, 643–646 (2006).
13. Geva-Zatorsky, N. *et al.* Oscillations and variability in the p53 system. *Mol. Syst. Biol.* **2**, 0033 (2006).
14. Kondo, T. *et al.* Circadian rhythms in prokaryotes: luciferase as a reporter of circadian gene expression in cyanobacteria. *Proc. Natl Acad. Sci. USA* **90**, 5672–5676 (1993).
15. Liu, Y., Golden, S. S., Kondo, T., Ishiura, M. & Johnson, C. H. Bacterial luciferase as a reporter of circadian gene expression in cyanobacteria. *J. Bacteriol.* **177**, 2080–2086 (1995).
16. Mihalcescu, I., Hsing, W. & Leibler, S. Resilient circadian oscillator revealed in individual cyanobacteria. *Nature* **430**, 81–85 (2004).
17. Andersen, J. B. *et al.* New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl. Environ. Microbiol.* **64**, 2240–2246 (1998).
18. Andersson, C. R. *et al.* Application of bioluminescence to the study of circadian rhythms in cyanobacteria. *Methods Enzymol.* **305**, 527–542 (2000).
19. Xu, Y., Mori, T. & Johnson, C. H. Cyanobacterial circadian clockwork: roles of KaiA, KaiB and the *kaiBC* promoter in regulating KaiC. *EMBO J.* **22**, 2117–2126 (2003).
20. Thattai, M. & van Oudenaarden, A. Intrinsic noise in gene regulatory networks. *Proc. Natl Acad. Sci. USA* **98**, 8614–8619 (2001).
21. Swain, P. S., Elowitz, M. B. & Siggia, E. D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl Acad. Sci. USA* **99**, 12795–12800 (2002).
22. Kepler, T. B. & Elston, T. C. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys. J.* **81**, 3116–3136 (2001).
23. Becskei, A., Kaufmann, B. B. & van Oudenaarden, A. Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nature Genet.* **37**, 937–944 (2005).
24. Volfson, D. *et al.* Origins of extrinsic variability in eukaryotic gene expression. *Nature* **439**, 861–864 (2006).
25. Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S. & Elowitz, M. B. Gene regulation at the single-cell level. *Science* **307**, 1962–1965 (2005).
26. Pedraza, J. M. & van Oudenaarden, A. Noise propagation in gene networks. *Science* **307**, 1965–1969 (2005).
27. Johnson, C. H. Global orchestration of gene expression by the biological clock of cyanobacteria. *Genome Biol.* **5**, 217 (2004).
28. Smith, R. M. & Williams, S. B. Circadian rhythms in gene transcription imparted by chromosome compaction in the cyanobacterium *Synechococcus elongatus*. *Proc. Natl Acad. Sci. USA* **103**, 8564–8569 (2006).
29. Vilar, J. M., Kueh, H. Y., Barkai, N. & Leibler, S. Mechanisms of noise-resistance in genetic oscillators. *Proc. Natl Acad. Sci. USA* **99**, 5988–5992 (2002).
30. Gonze, D., Halloy, J. & Goldbeter, A. Robustness of circadian rhythms with respect to molecular noise. *Proc. Natl Acad. Sci. USA* **99**, 673–678 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. S. Golden and J. L. Ditty for assistance with the initial phase of this work and their gifts of plasmids and strains. We acknowledge I. Lipchin and M. J. T. O'Kelly for assistance with data collection, cloning and bioluminescence measurements. We acknowledge A. Tolonen, S. W. Chisholm, M. Thattai, H. Lim, J. C. Gore and A. Raj for discussions and suggestions. This work was performed in part at the MIT Laser Biomedical Research Center. This work was supported by NSF and NIH grants.

Author Contributions J.R.C. and P.L. performed the experiments. J.M.P. developed the model. J.R.C., .J.M.P. and A.v.O. designed the experiments, interpreted the results and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.v.O. (avano@mit.edu).

LETTERS

Toll-like receptor signalling in macrophages links the autophagy pathway to phagocytosis

Miguel A. Sanjuan[1], Christopher P. Dillon[1], Stephen W. G. Tait[1], Simon Moshiach[2], Frank Dorsey[3], Samuel Connell[1], Masaaki Komatsu[4], Keiji Tanaka[4], John L. Cleveland[5], Sebo Withoff[1] & Douglas R. Green[1]

Phagocytosis and autophagy are two ancient, highly conserved processes involved, respectively, in the removal of extracellular organisms and the destruction of organisms in the cytosol[1–3]. Autophagy, for either metabolic regulation or defence, involves the formation of a double membrane called the autophagosome, which then fuses with lysosomes to degrade the contents[4], a process that has similarities with phagosome maturation. Toll-like-receptor (TLR) engagement activates a variety of defence mechanisms within phagocytes[5], including facilitation of phagosome maturation[6], and also engages autophagy[7]. Therefore we speculated that TLR signalling might link these processes to enhance the function of conventional phagosomes. Here we show that a particle that engages TLRs on a murine macrophage while it is phagocytosed triggers the autophagosome marker LC3 to be rapidly recruited to the phagosome in a manner that depends on the autophagy pathway proteins ATG5 and ATG7; this process is preceded by recruitment of beclin 1 and phosphoinositide-3-OH kinase activity. Translocation of beclin 1 and LC3 to the phagosome was not associated with observable double-membrane structures characteristic of conventional autophagosomes, but was associated with phagosome fusion with lysosomes, leading to rapid acidification and enhanced killing of the ingested organism.

A critical step in the autophagy process is the ligation of LC3 to phosphatidylethanolamine, resulting in aggregates of LC3[8]. In order to visualize processes related to autophagy, we used green fluorescent protein (GFP)–LC3[9] in the RAW macrophage cell line (RAW-GFP–LC3). As shown previously for lipopolysaccharide (LPS)[7], on stimulation of TLRs with LPS, CpG, or imiquimod for several hours, we observed the formation of LC3 aggregates (Supplementary Fig. 1a; quantification is shown in Supplementary Fig. 1b) and the generation of the conjugated form of LC3 (Supplementary Fig. 1c). We then examined the effects of TLR engagement on LC3 localization during phagocytosis. Latex beads with or without the associated TLR ligand PAM3CSK4 were fed to RAW-GFP–LC3 cells (Fig. 1a). The presence of the TLR agonist induced a rapid recruitment of LC3 to the phagosome, which was not seen in the absence of the agonist. Similar effects were observed on ingestion of beads with LPS (Fig. 1a), killed yeast (zymosan) (Fig. 1b, time course shown in Supplementary Fig. 2a) or *Escherichia coli* bacteria (Supplementary Fig. 3a, 3b, and Supplementary Movie 1). Internalization of particles was confirmed after each experiment by visual microscopic inspection (Supplementary Fig. 2b). Time lapse analysis of the recruitment of GFP–LC3 to phagosomes (Fig. 1c and Supplementary Movie 2) revealed that this process occurred rapidly (≤15 min after internalization) and was transient, with the effect waning after approximately 60 min.

Zymosan is recognized by TLR2 (ref. 10) and therefore we used mice deficient in this receptor[11] to test the role of TLR in LC3 recruitment to the phagosome. Primary macrophages from *Tlr2*$^{-/-}$ mice showed a reduced translocation of LC3 in response to zymosan (Fig. 1d, quantified in Fig. 1e). This was consistent with the reduced (but not fully deficient) TLR signalling described for zymosan in the absence of TLR2 (ref. 12). In another approach, 293T cells transduced with GFP–LC3 and transiently transfected with TLR2 were cultured with zymosan (Supplementary Fig. 3c). Although phagocytosis by these cells is inefficient, phagosomes containing zymosan were observed to recruit GFP–LC3 only in those cells expressing TLR2. Thus, TLR signalling during phagocytosis induces a rapid recruitment of LC3 to the phagosome. However, we found that addition of the TLR ligands LPS or PAM3CSK4 to RAW cells that had engulfed latex beads did not induce recruitment of LC3 to the phagosomes at any time point (data not shown), suggesting that localized TLR signalling within the phagosome may be important for the recruitment process.

The formation of double-membrane structures and autophagosomes[8] were readily observed by electron microscopy of cells treated with rapamycin to induce autophagy or treatment with chloroquine to cause accumulation of autophagosomes[13] (Supplementary Fig. 4a). However, we failed to observe formation of such double-membrane structures associated with phagocytosed zymosan (Fig. 2a) despite recruitment of LC3. Pre-treatment of cells to induce autophagy (rapamycin or starvation) or the accumulation of autophagic vesicles (chloroquine) did not induce recruitment of LC3 to phagosomes in the absence of TLR signalling, or affect this recruitment in the presence of such signalling (Fig. 2b and Supplementary Fig. 5); furthermore, it did not cause the appearance of double-membrane structures associated with phagocytosed zymosan (Fig. 2a).

The autophagic vacuole has been suggested to derive from the endoplasmic reticulum (ER), based on localization of luminal ER proteins[14] and a requirement for Sec proteins in autophagosome formation[15]. In RAW cells treated with chloroquine, the dye ER-tracker co-localized with GFP–LC3 (Supplementary Fig. 4b and Supplementary Movie 3), representing autophagy of ER components[16] or a contribution of the ER to the autophagosome. No such co-localization was observed when cells engulfed zymosan, inducing GFP–LC3 association with the phagosome (Supplementary Fig. 4c and Supplementary Movie 4), again possibly distinguishing this phenomenon from autophagy.

Autophagosome formation requires a number of components of the autophagy pathway. Knockdown of ATG5 (Supplementary Fig. 6) markedly reduced the recruitment of GFP–LC3 to the zymosan-containing phagosomes (Fig. 3a, b). The role of the autophagy

[1]Department of Immunology, [2]Department of Tumor Cell Biology, and [3]Department of Biochemistry, St Jude Children's Research Institute, Memphis, Tennessee 38105, USA. [4]Laboratory of Frontier Science, Tokyo Metropolitan Institute of Medical Science, Bunkyo-ku, Tokyo 113-8613, Japan. [5]Department of Cancer Biology, The Scripps Research Institute, Jupiter, Florida 33458, USA.

pathway was further tested using macrophages derived from wild-type versus *Atg7*$^{-/-}$ mice. Whereas primary macrophages displayed association of GFP–LC3 with the phagosome after uptake of zymosan, no such association was observed in ATG7-deficient cells (Fig. 3c, d).

The initiation of the autophagy pathway involves members of the phosphoinositide-3-OH kinase (PI(3)K) family[17]. TLRs can similarly activate PI(3)K[18]. Pharmacological inhibition of PI(3)K using wortmannin or LY294002 (LY29) inhibited GFP–LC3 translocation (Fig. 3e, quantified in Fig. 3f). Wortmannin was added approximately 10 min after zymosan internalization to minimize the effects on engulfment[19]. Because phagocytosis itself engages PI(3)K[20], we used a probe derived from a phosphatidylinositol trisphosphate (PtdIns3P)-binding PX-domain of p40 (phox)[21] to examine the appearance of PtdIns3P on the phagosome with or without TLR signalling. Transient, low-level PtdIns3P was seen in the absence of TLR signalling, whereas more persistent, higher levels were observed with TLR signalling (Fig. 3g, Supplementary Fig. 7, and Supplementary Movie 5), beginning before association of GFP–LC3 (Fig. 3g, Supplementary Fig. 7 and Supplementary Movies 6 and 7).

Initiation of the autophagy pathway involves beclin 1 (BECN1) in a complex with the PI(3)K hVps34 (also known as PIK3C3; ref. 4). Although no association of GFP–BECN1 with phagosomes occurred with latex beads (Supplementary Movie 8), zymosan induced an almost instantaneous association of GFP–BECN1 (as soon as 1 min after internalization) with the phagosome (Fig. 3h and Supplementary Movie 9).

TLR signalling proceeds, in part, through the adaptor molecule MyD88 (ref. 22). We found, however, that the absence of MyD88 did not affect the ability of TLR signalling by zymosan (Supplementary Fig. 8a) or red fluorescent *E. coli* (Supplementary Fig. 8b) to induce

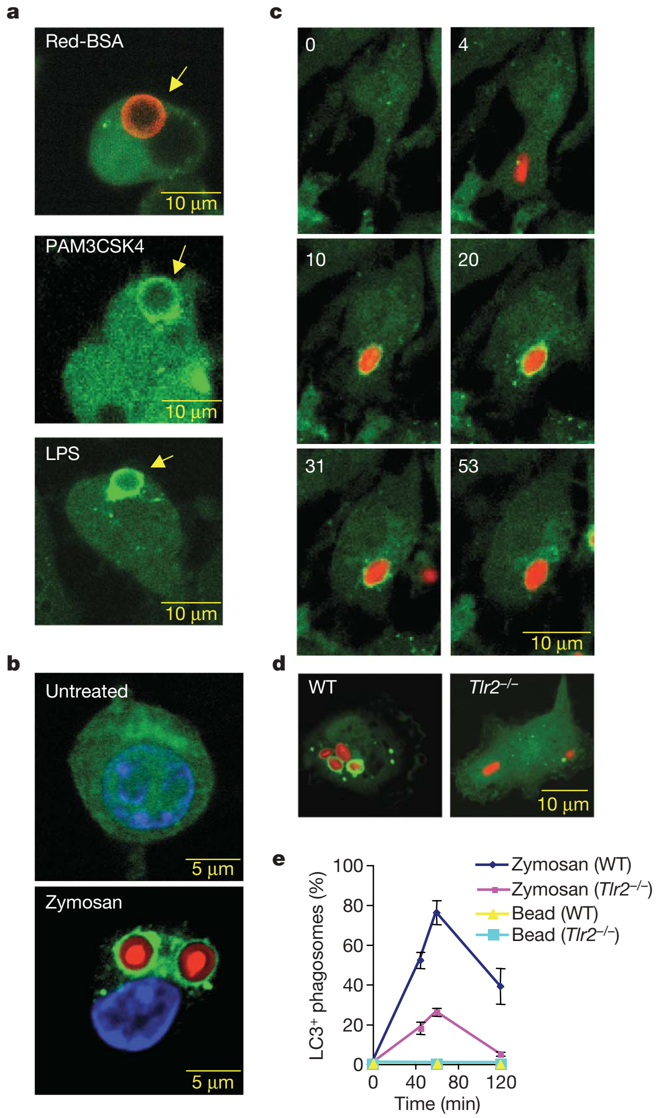


Figure 1 | TLR signalling activates components of the autophagy pathway. **a**, Internalization of beads with PAM3CSK, LPS, or Alexa-Fluor-555-labelled BSA (Red-BSA; yellow arrows) and association with GFP–LC3 in RAW cells was followed by time-lapse video for 3 h, 1 frame/2 min (representative frames are shown). **b**, GFP–LC3 in RAW cells translocates to phagosomes containing zymosan (red), assessed as in **a**. **c**, Time course of translocation of GFP–LC3 to a phagosome containing zymosan (red) (see Supplementary Movie 2). Time (minutes) is indicated in the top left of each panel. **d**, LC3 translocation to the phagosome is induced by TLR signalling. GFP–LC3 was transiently transfected into wild-type or *Tlr2*$^{-/-}$ macrophages, which were exposed to zymosan for 45 min. **e**, The percentage of phagosomes ($n \geq 50$ per group) with GFP–LC3 in wild-type and *Tlr2*$^{-/-}$ macrophages. Means and ranges are shown (4 mice per group were used).

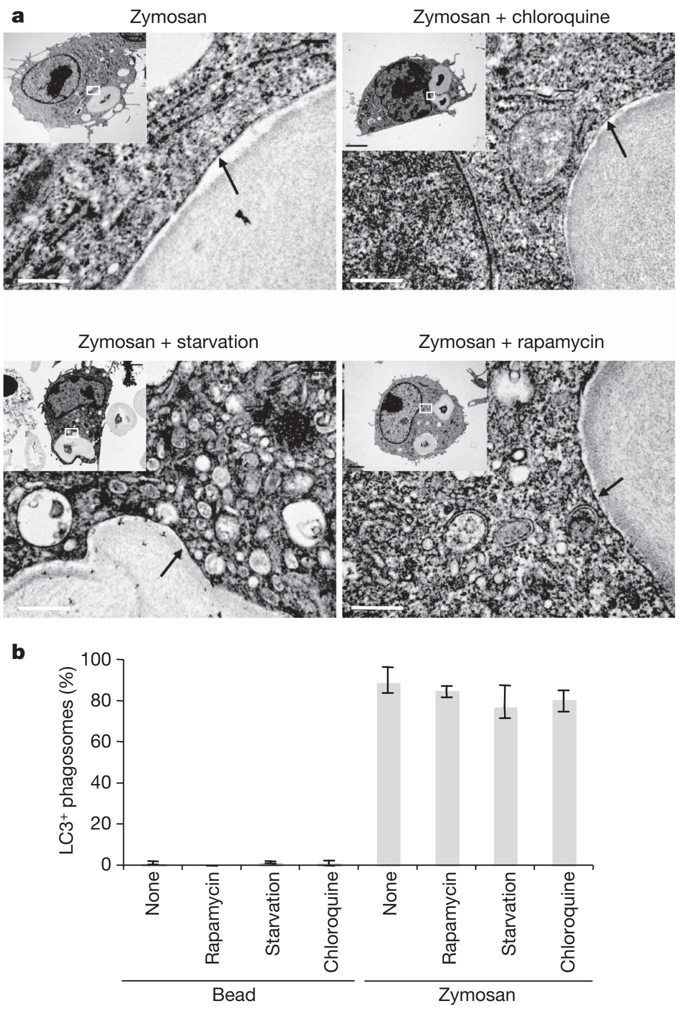


Figure 2 | TLR-induced LC3 localization to the phagosome appears to occur independently of autophagy. **a**, RAW-GFP–LC3 cells were treated with chloroquine (50 μM, 3 h) or rapamycin (200 nM, 6 h), or starved (6 h). Subsequently, the cells were exposed to zymosan for 1 h. We did not detect any double membranes associated with the phagocytosed zymosan (black arrows highlight single membranes) (scale bars equal 0.4 μm). **b**, RAW-GFP–LC3 cells were pre-treated with chloroquine (50 μM), starvation media, or rapamycin (200 nM) for 3 h. Cells were then fed with zymosan or beads and followed by time-lapse video at 3-min intervals for 3 h. All LC3-associated phagosomes ($n \geq 100$ per group) at all time points were quantified in three independent experiments (means and ranges are shown).

association of GFP–LC3 with the phagosome in primary macrophages. Pharmacological inhibition of p38, also implicated in TLR signalling[23], had a partial effect on translocation of GFP–LC3, reducing the frequency of associated phagosomes by about 30% (Supplementary Fig. 8c).

During phagocytosis of beads, limited fusion of the phagosomes with lysosomes (labelled with lysotracker red) was observed (Fig. 4a, b and Supplementary Movie 10). In contrast, beads coupled with the TLR ligand PAM3CSK4 induced a more rapid and extensive acidification (Fig. 4a, b, and Supplementary Movie 11). Similarly, phagosomes that had engulfed zymosan particles bound GFP–LC3 and then fused with lysosomes (Fig. 4c). This effect was inhibited by knockdown of ATG5 (Fig. 4d). These results suggest that TLR-induced association of elements of the classical autophagy pathway to the phagosome promotes fusion with lysosomes.

At 4 h after phagocytosis, live *Saccharomyces cerevisiae* engulfed by ATG7-deficient macrophages showed a marked survival advantage over yeast taken up by wild-type macrophages (Fig. 4e). This strongly suggests that the TLR-induced recruitment of elements of the classical autophagy pathway to the phagosome has an impact on survival of engulfed organisms, especially at early time points. No differences in ATP levels in wild-type or *Atg7*$^{-/-}$ macrophages were detected (data not shown), and therefore the effects were not due to energetic defects in the null cells.

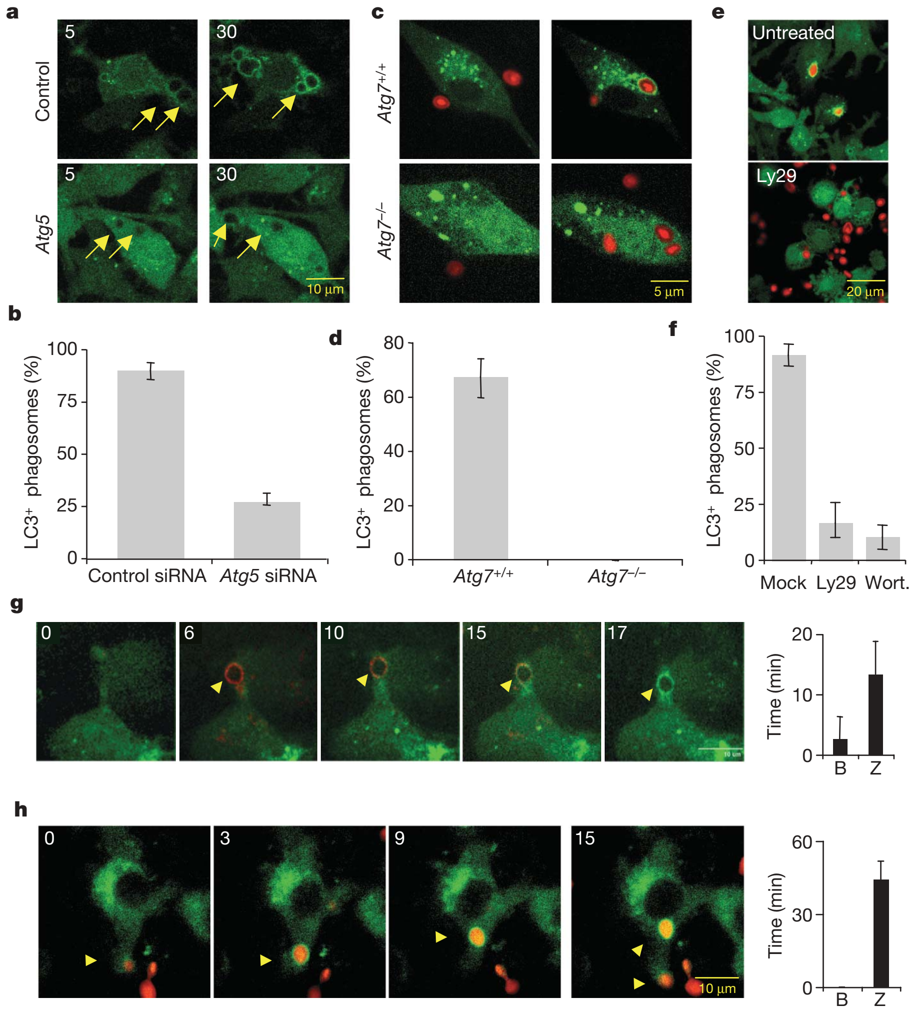


Figure 3 | Components of the autophagic pathway induce localization of LC3 to the phagosome on TLR stimulation. a, RAW cells expressing GFP–LC3 were transfected by electroporation with control or *Atg5* siRNA oligonucleotides. At 24 h after transfection, cells were fed with zymosan (yellow arrows) and the internalization process was followed by time-lapse video microscopy. Images obtained at 5 and 30 min after zymosan engulfment are shown. **b**, The percentage of GFP–LC3-associated phagosomes ($n \geq 100$ per /group) was obtained from three independent time-lapse videos (2 h each) of macrophages transfected with control or *Atg5* siRNA (means and ranges are shown). **c**, Bone-marrow-derived macrophages from wild-type and *Atg7*$^{-/-}$ mice were transiently transfected with GFP–LC3. After 24 h macrophages were fed with zymosan (red) for 1 h. Confocal images of the same cell before and after zymosan engulfment are shown. **d**, The percentage of GFP–LC3-associated phagosomes ($n \geq 50$ per group) was assessed after 1.5 h in wild-type and *Atg7*$^{-/-}$ macrophages (represented as means and ranges; three mice per group). **e**, Ly294002 (Ly29, 10 µM) was added 10 min before zymosan, and wortmannin was added after engulfment but before GFP–LC3 translocation in RAW-GFP–LC3 cells. Representative images are shown. **f**, Percentage of GFP–LC3-associated phagosomes ($n \geq 50$ per group) with P(I)3K inhibitors assessed by time-lapse video (means and ranges are shown; three independent 3 h videos). **g**, RAW cells expressing GFP–LC3 and the PX domain of p40(phox) fused to mCherry (red) were fed with zymosan (yellow arrows) or with beads (Supplementary Fig. 6). Internalization was followed at 1-min intervals (Supplementary Movie 5), representative images are shown. Single cells were followed for at least 2 h ($n \geq 20$ per group) (Supplementary Movies 6, 7 and Supplementary Fig. 6). Mean ± s.d. duration of mCherry-PX translocation to the phagosome for beads (B) and zymosan (Z) are shown. **h**, Internalization of zymosan (yellow arrows) was followed at 1.5-min intervals for at least 2 h in RAW cells transiently transfected with GFP–BECN1. Representative images are shown. Translocation of GFP–BECN1 to the phagosome for beads (B) and zymosan (Z) was quantified (mean ± s.d.) using time-lapse movies (Supplementary Movie 8 and 9; $n \geq 20$ cells per group). For **g** and **h**, time (in minutes) is indicated at the top left of each panel.

Several recent studies have implicated autophagy in the removal of pathogens located in phagosomes, including *Mycobacterium tuberculosis*[3] and *Toxoplasma gondii*[24], as well as those that invade the cytosol, such as *Shigella flexneri*[25] and group A *Streptococcus*[2]. Autophagosomes either enclose the pathogens or fuse with the phagosome, and in at least one case this promotes phagosome maturation[24].

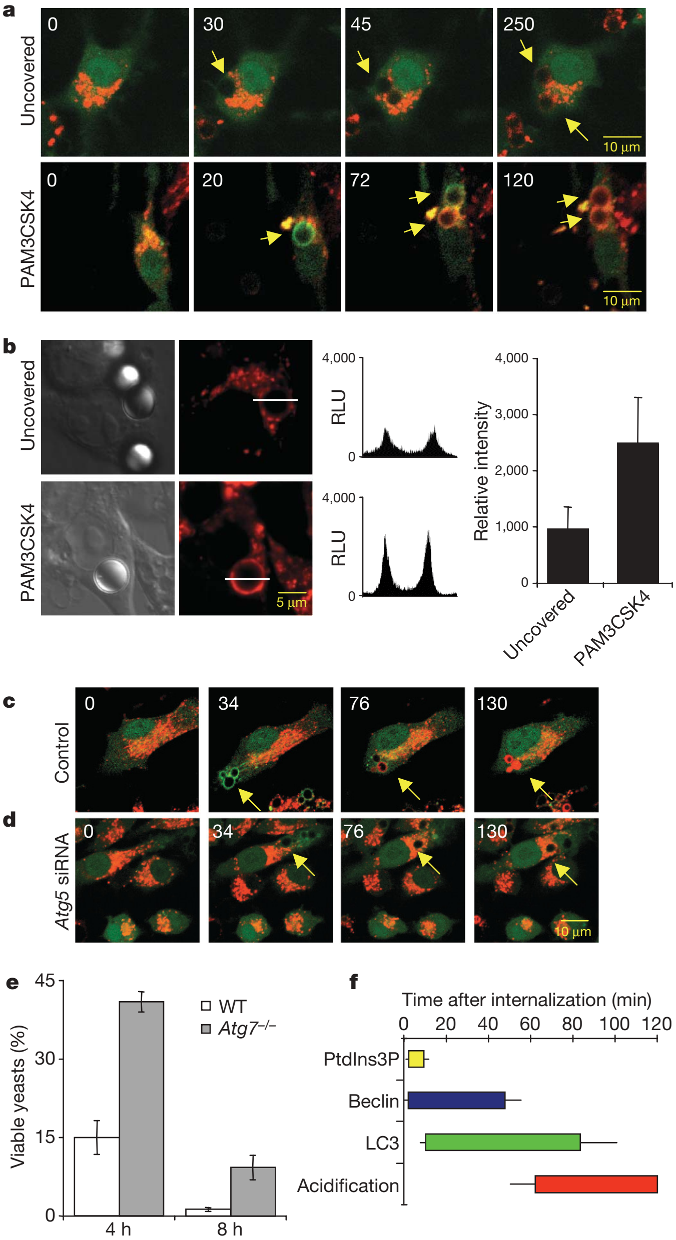


Figure 4 | TLR signalling induces phagosome maturation using components of the autophagic pathway. **a**, Internalization of control or PAM3CSK4-covered beads was followed in RAW-GFP–LC3 cells pre-loaded with lysotracker (red) by time-lapse video for 3 h. Time (in minutes) is indicated at the top left of each panel. **b**, RAW cells pre-loaded with lysotracker were exposed to control or PAM3CSK-covered beads for 6 h. Acidification of the beads ($n \geq 20$ per group), measured as maximum lysotracker relative light units (RLU) on a horizontal cross-section of each bead represented in the x axis (mean ± s.d.; Wilcoxon–Mann–Whitney test, $P < 0.0005$), is shown. **c**, **d**, Knockdown of ATG5 reduced zymosan-induced GFP–LC3 translocation and lysosomal fusion. **e**, Primary macrophages from wild-type (white bars) and *Atg7*-null mice (grey bars) were fed live yeast for 1 h and washed. Internalized yeast were extracted from the macrophages at 4 h and 8 h after re-plating. Means and ranges from six different measurements using two mice per group are shown. **f**, Initiation and duration of translocation of PX-mCherry to phagosomes (yellow bar), GFP–BECN1 translocation (blue bar), GFP–LC3 translocation (green bar) and Lysotracker recruitment (acidification) were determined during the first 2 h after internalization (10 cells per group, error bars indicate s.d.).

A recent publication[7] shows that engagement of TLR4 enhances conventional autophagy and autophagy directed to infectious organisms in the cytosol. We have found, in contrast, that TLR signalling during phagocytosis of an extracellular organism usurps the autophagy pathway to associate LC3 rapidly with the phagosome, apparently without the formation of conventional autophagosomes in this time period. We cannot, however, formally exclude that the association of LC3 with the phagosome is due to widespread and extremely rapid formation of small autophagosomes that fuse with the phagosome. Alternatively, vesicles derived from the secretory pathway and associated with elements of the autophagy pathway (including LC3) may dock, flatten and fuse with the phagosome, as described for phagosomes containing *Legionella*[26,27]. However, we were unable to observe this by either electron microscopy or in live cell imaging, even when autophagosomes were increased in the cells by rapamycin treatment or starvation to induce autophagy, or through the inhibitory action of chloroquine. Furthermore, the extremely rapid association of the phagosome with BECN1 (approximately 1 min) and a strong PI(3)K activity, followed by association with LC3 only a few minutes later, suggests the possibility that elements of the classical autophagy pathway are effectively and efficiently recruited to the phagosome on TLR signalling.

Although some studies[6,28] have shown that TLR signalling by means of MyD88 can enhance phagosome acidification and function, another study showed no effect of TLR signalling after engulfment of opsonized particles[29,30]. Here, we did not find a requisite role for MyD88 in the induction of translocation of LC3 to the phagosome, but found that engaging the autophagy pathway via TLR signalling enhanced phagosome maturation and destruction of the engulfed yeast.

METHODS SUMMARY

Mice and cells. All mice were bred and housed pathogen-free in our facility. Mouse primary macrophages were differentiated from fetal liver or from bone marrow and transiently transfected with GFP–LC3. RAW cells were stably transfected with GFP–LC3 (RAW-GFP–LC3) with or without PX-mCherry (RAW-LC3/Px), or transiently with BECN1–GFP.

Phagocytosis. Beads used for internalization were prepared by conjugating Alexa-Fluor-555-labelled BSA (Invitrogen) (Red-BSA) or PAM3CSK4 (Invivogen) to 6-µm diameter latex microspheres. Sonicated biotinylated-LPS (0.5 mg ml^{-1}) (Invivogen) was conjugated to SPHERO Streptavidin-polystyrene beads (Lybertiville). Zymosan particles were added at a ratio of 8:1 (particle:cell), whereas beads were added at a ratio of 10:1 (bead:cell). At the end of each experiment, we confirmed that particles were completely internalized by differential focusing.

Cell imaging. Two systems were used for live cell imaging: Marianas spinning-disk confocal imaging system (Intelligent Imaging Innovations/3i) and laser scanning confocal microscopy using a C1Si confocal system (Nikon). Laser scanning confocal microscopy on fixed tissues was performed with a Zeiss Axioplan 2 upright microscope equipped with an LSM 510 NLO confocal system (Carl Zeiss MicroImaging). For time-lapse microscopic imaging of phagocytosis, macrophages were plated on collagen-coated glass-bottom dishes (MatTek). Translocation of probes to the phagosome was quantified in time lapse movies.

Received 25 September; accepted 22 October 2007.

1. Levine, B. Eating oneself and uninvited guests: autophagy-related pathways in cellular defense. *Cell* **120**, 159–162 (2005).
2. Nakagawa, I. *et al.* Autophagy defends cells against invading group A *Streptococcus*. *Science* **306**, 1037–1040 (2004).
3. Singh, S. B., Davis, A. S., Taylor, G. A. & Deretic, V. Human IRGM induces autophagy to eliminate intracellular mycobacteria. *Science* **313**, 1438–1441 (2006).
4. Levine, B. & Deretic, V. Unveiling the roles of autophagy in innate and adaptive immunity. *Nature Rev. Immunol.* **7**, 767–777 (2007).
5. Takeuchi, O. & Akira, S. Signaling pathways activated by microorganisms. *Curr. Opin. Cell Biol.* **19**, 185–191 (2007).

6. Blander, J. M. & Medzhitov, R. Regulation of phagosome maturation by signals from toll-like receptors. *Science* **304,** 1014–1018 (2004).
7. Xu, Y. *et al.* Toll-like receptor 4 is a sensor for autophagy associated with innate immunity. *Immunity* **27,** 135–144 (2007).
8. Tanida, I., Ueno, T. & Kominami, E. LC3 conjugation system in mammalian autophagy. *Int. J. Biochem. Cell Biol.* **36,** 2503–2518 (2004).
9. Mizushima, N., Yamamoto, A., Matsui, M., Yoshimori, T. & Ohsumi, Y. *In vivo* analysis of autophagy in response to nutrient starvation using transgenic mice expressing a fluorescent autophagosome marker. *Mol. Biol. Cell* **15,** 1101–1111 (2004).
10. Ozinsky, A. *et al.* The repertoire for pattern recognition of pathogens by the innate immune system is defined by cooperation between toll-like receptors. *Proc. Natl Acad. Sci. USA* **97,** 13766–13771 (2000).
11. Wooten, R. M. *et al.* Toll-like receptor 2 is required for innate, but not acquired, host defense to *Borrelia burgdorferi*. *J. Immunol.* **168,** 348–355 (2002).
12. Brown, G. D. *et al.* Dectin-1 is a major beta-glucan receptor on macrophages. *J. Exp. Med.* **196,** 407–412 (2002).
13. Schmid, D., Pypaert, M. & Munz, C. Antigen-loading compartments for major histocompatibility complex class II molecules continuously receive input from autophagosomes. *Immunity* **26,** 79–92 (2007).
14. Dunn, W. A. Jr. Studies on the mechanisms of autophagy: formation of the autophagic vacuole. *J. Cell Biol.* **110,** 1923–1933 (1990).
15. Ishihara, N. *et al.* Autophagosome requires specific early Sec proteins for its formation and NSF/SNARE for vacuolar fusion. *Mol. Biol. Cell* **12,** 3690–3702 (2001).
16. Bernales, S., Schuck, S. & Walter, P. ER-Phagy: selective autophagy of the endoplasmic reticulum. *Autophagy* **3,** 285–287 (2007).
17. Yorimitsu, T. & Klionsky, D. J. Autophagy: molecular machinery for self-eating. *Cell Death Differ.* **12** (suppl. 2), 1542–1552 (2005).
18. Kuo, C. C., Lin, W. T., Liang, C. M. & Liang, S. M. Class I and III phosphatidylinositol 3′-kinase play distinct roles in TLR signaling pathway. *J. Immunol.* **176,** 5943–5949 (2006).
19. Gillooly, D. J., Simonsen, A. & Stenmark, H. Phosphoinositides and phagocytosis. *J. Cell Biol.* **155,** 15–17 (2001).
20. Leverrier, Y. *et al.* Class I phosphoinositide 3-kinase p110β is required for apoptotic cell and Fcγ receptor-mediated phagocytosis by macrophages. *J. Biol. Chem.* **278,** 38437–38442 (2003).
21. Kanai, F. *et al.* The PX domains of p47phox and p40phox bind to lipid products of PI(3)K. *Nature Cell Biol.* **3,** 675–678 (2001).
22. Janssens, S. & Beyaert, R. A universal role for MyD88 in TLR/IL-1R-mediated signaling. *Trends Biochem. Sci.* **27,** 474–482 (2002).
23. Vasselon, T., Hanlon, W. A., Wright, S. D. & Detmers, P. A. Toll-like receptor 2 (TLR2) mediates activation of stress-activated MAP kinase p38. *J. Leukoc. Biol.* **71,** 503–510 (2002).
24. Andrade, R. M., Wessendarp, M., Gubbels, M. J., Striepen, B. & Subauste, C. S. CD40 induces macrophage anti-*Toxoplasma gondii* activity by triggering autophagy-dependent fusion of pathogen-containing vacuoles and lysosomes. *J. Clin. Invest.* **116,** 2366–2377 (2006).
25. Ogawa, M. *et al.* Escape of intracellular *Shigella* from autophagy. *Science* **307,** 727–731 (2005).
26. Amer, A. O. & Swanson, M. S. Autophagy is an immediate macrophage response to *Legionella pneumophila*. *Cell. Microbiol.* **7,** 765–778 (2005).
27. Tilney, L. G., Harb, O. S., Connelly, P. S., Robinson, C. G. & Roy, C. R. How the parasitic bacterium *Legionella pneumophila* modifies its phagosome and transforms it into rough ER: implications for conversion of plasma membrane to the ER membrane. *J. Cell Sci.* **114,** 4637–4650 (2001).
28. Blander, J. M. & Medzhitov, R. Reply to ''Toll-like receptors and phagosome maturation''. *Nature Immunol.* **8,** 217–218, doi:10.1038/ni0307-217b (2007).
29. Yates, R. M. & Russell, D. G. Phagosome maturation proceeds independently of stimulation of toll-like receptors 2 and 4. *Immunity* **23,** 409–417 (2005).
30. Russell, D. G. & Yates, R. M. Toll-like receptors and phagosome maturation. *Nature Immunol.* **8,** 217, doi:10.1038/ni0307-217a (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Murray, M. del Mar Lozano-Cedeño, C.-S. Li, M. Smeltzer, M. Bix and D. Vignali for discussions, and the St Jude Flow Cytometry and Scientific Imaging shared resource facilities for technical assistance. M. Colonna and H. Haecker provided us with essential reagents. This work is supported by grants from the US National Institutes of Health.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.R.G. (douglas.green@stjude.org).

Vol 450|20/27 December 2007|doi:10.1038/nature06388

LETTERS

Cdc48/p97 promotes reformation of the nucleus by extracting the kinase Aurora B from chromatin

Kristijan Ramadan[1]*, Roland Bruderer[1]*, Fabio M. Spiga[1,2], Oliver Popp[1], Tina Baur[1], Monica Gotta[1,2] & Hemmo H. Meyer[1]

During division of metazoan cells, the nucleus disassembles to allow chromosome segregation, and then reforms in each daughter cell. Reformation of the nucleus involves chromatin decondensation and assembly of the double-membrane nuclear envelope around the chromatin; however, regulation of the process is still poorly understood[1,2]. *In vitro*, nucleus formation requires p97 (ref. 3), a hexameric ATPase implicated in membrane fusion and ubiquitin-dependent processes[4,5]. However, the role and relevance of p97 in nucleus formation have remained controversial. Here we show that p97 stimulates nucleus reformation by inactivating the chromatin-associated kinase Aurora B. During mitosis, Aurora B inhibits nucleus reformation by preventing chromosome decondensation and formation of the nuclear envelope membrane. During exit from mitosis, p97 binds to Aurora B after its ubiquitylation and extracts it from chromatin. This leads to inactivation of Aurora B on chromatin, thus allowing chromatin decondensation and nuclear envelope formation. These data reveal an essential pathway that regulates reformation of the nucleus after mitosis and defines ubiquitin-dependent protein extraction as a common mechanism of Cdc48/p97 activity also during nucleus formation.

The abundant and conserved molecular chaperone p97 (Cdc48 in yeast, CDC-48 in *Caenorhabditis elegans* and valosin-containing protein (VCP) in humans), a member of the AAA (ATPases associated with various cellular activities) protein family, is involved in many different cellular processes ranging from bulk degradation of endoplasmic reticulum (ER) proteins to cell-cycle regulation and organelle biogenesis[4,5]. With its heterodimeric Ufd1–Npl4 cofactor, it is essential for the formation of a closed nuclear envelope around decondensed chromatin in *Xenopus laevis* egg extract[3], suggesting an important role in nucleus formation.

Two models may explain the role of p97 in the process. First, p97 may directly mediate the required membrane fusion, as has been suggested for this and other processes[2,6,7]. We favoured a second model, in which p97 regulates nucleus formation by removing an inhibitor from the forming nucleus. This notion is based on analogy to its function in ER-related processes in interphase, where p97 binds via Ufd1–Npl4 to ubiquitylated substrate proteins in the membrane. Upon ATP hydrolysis, it extracts them into the cytosol, where they are either degraded by the proteasome or de-ubiquitylated[4,5,8].

Indirect evidence suggests that this p97-regulated inhibitor might be Aurora B, a mitotic kinase that, in complex with other proteins including survivin and INCENP, regulates and coordinates different aspects of mitosis including chromosome segregation and cytokinesis[9]. First, $p97^{Ufd1–Npl4}$ physically interacts with survivin in *Xenopus* egg extracts[10]. Second, Aurora B catalyses processes that may antagonize nucleus formation. It contributes to chromatin condensation during mitosis[9,11–14], and phosphorylates histone H3 at serine 10, which indirectly may regulate recruitment of nuclear envelope membrane to chromatin[15–18].

We first tested whether Aurora B inhibits formation of the nuclear envelope around decondensed sperm chromatin in *Xenopus* interphase egg extract. Addition of recombinant wild-type (WT) Aurora B[19], but not a kinase-inactive form (K122R) nor the WT in combination with the Aurora B inhibitor hesperadin[20], prevented nuclear envelope formation (Fig. 1a and Supplementary Fig. 1a). Moreover, addition of okadaic acid, which activates Aurora B and inhibits dephosphorylation of its substrates[15,21] and those of other kinases, blocked nuclear envelope formation and even led to chromatin condensation (Supplementary Fig. 1b). Crucially, nucleus formation was restored by inactivation of Aurora B with hesperadin (Supplementary Fig. 1b). We next asked whether Aurora B played a role in $p97^{Ufd1–Npl4}$-mediated regulation of nuclear envelope formation. As expected from previous experiments[3], addition of the dominant-negative variant of p97, p97ΔD2 (ref. 22), or a monoclonal antibody against the p97 cofactor Ufd1 inhibited nuclear envelope formation (Fig. 1c, d). By contrast, immunodepletion of Aurora B to non-detectable levels did not affect nuclear envelope formation (Fig. 1b–d). Importantly, however, removal of Aurora B rendered nuclear envelope formation largely insensitive to inhibition of $p97^{Ufd1–Npl4}$ (Fig. 1c, d). To confirm this observation, we depleted p97, which again led to inhibition of nuclear envelope formation (Fig. 1e). Consistently, inactivation of Aurora B by hesperadin again restored nuclear envelope formation (Fig. 1e), whereas the inhibitor alone had no effect. Thus, $p97^{Ufd1–Npl4}$ is largely dispensable for nuclear envelope formation in the absence of active Aurora B. This strongly suggests that $p97^{Ufd1–Npl4}$ is not directly involved in any essential mechanism of membrane fusion and supports the notion that $p97^{Ufd1–Npl4}$ functions as a regulator that restrains the inhibitory activity of Aurora B.

We next investigated whether this regulation may be direct. Indeed, immunoprecipitation of Aurora B from egg extract revealed that p97 physically interacts with the kinase (Fig. 2a). We also confirmed p97 binding to survivin[10], but could not detect any specific interaction with INCENP (Fig. 2a). In other processes, p97 binds to substrates after their ubiquitylation[4,5,8]. Indeed, fractions of Aurora B and survivin, but not INCENP, were ubiquitylated with lysine-48-linked ubiquitin chains in egg extract (Fig. 2b). In contrast to what was proposed for survivin in human somatic cells[10], neither survivin nor Aurora B was modified with lysine-63-linked ubiquitin in egg extract (Fig. 2b). To clarify whether $p97^{Ufd1–Npl4}$ interacts with the ubiquitylated or the non-modified form of Aurora B, we immunoisolated the $p97^{Ufd1–Npl4}$ complex with monoclonal antibodies

[1]Institute of Biochemistry, ETH Zurich, 8093 Zurich, Switzerland. [2]Department of Genetic Medicine and Development, University of Geneva School of Medicine, Rue Michel-Servet 1, 1211 Geneva 4, Switzerland.
*These authors contributed equally to this work.

against Ufd1. The isolate did not contain any non-modified Aurora B above background (Fig. 2c), but a large amount of ubiquitin-conjugates that most likely represented a mixture of substrates. We released these p97$^{\mathrm{Ufd1-Npl4}}$-associated proteins by denaturation in detergent. Subsequent precipitation with antibodies against Aurora B specifically isolated ubiquitin conjugates, demonstrating that Aurora B is associated with p97$^{\mathrm{Ufd1-Npl4}}$ in a ubiquitylated form and suggesting that it is a substrate of p97 (Fig. 2c).

Although Aurora B is ubiquitylated in the egg extract, it is not quantitatively degraded in the first embryonic cell cycles and was stable in our extract (Supplementary Fig. 2). This suggests that ubiquitylation and p97 activity do not mediate bulk degradation of Aurora B, but rather regulate its activity or localization. Indeed, we observed an increase of chromatin-associated Aurora B upon incubation in extract with p97ΔD2, compared with control reactions with buffer alone or the inactive double-mutant p97ΔD2(K251A)[22] (Supplementary Fig. 3). Importantly, incubation of chromatin with p97ΔD2 led to a more than twofold increase of H3 kinase activity on chromatin, but not in cytosol (Fig. 2d), indicating that p97

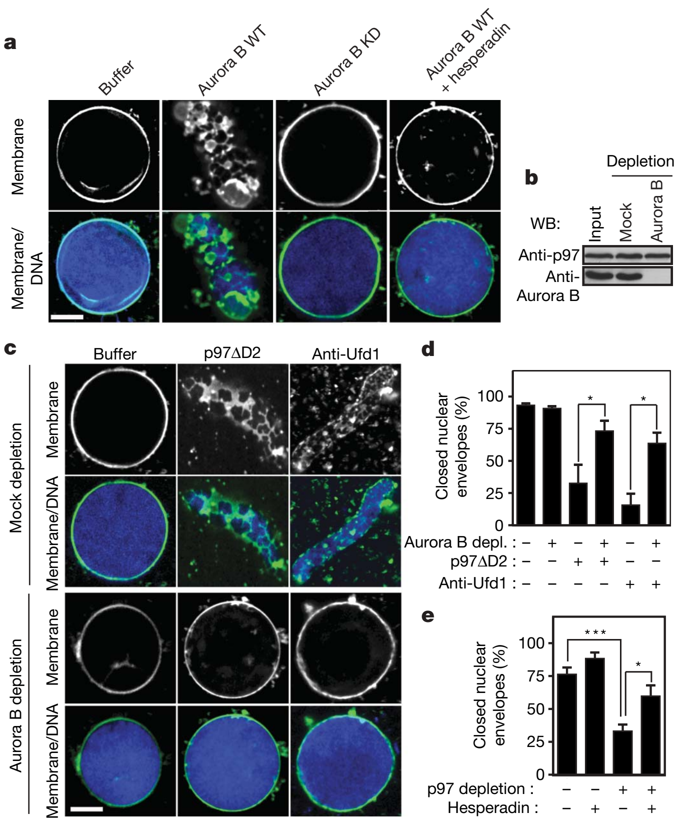


Figure 1 | p97 promotes nuclear envelope formation by antagonizing the inhibitory activity of Aurora B kinase. **a**, Sperm chromatin was incubated with *Xenopus* interphase egg extract with or without recombinant Aurora B (3.4 μM) or a kinase-dead (KD) variant in the presence or absence of the Aurora B inhibitor hesperadin (1 μM). Chromatin was fixed, and stained with 4,6-diamidino-2-phenylindole (DAPI) and the membrane dye $DiOC_6$. Scale bar, 5 μM. See Supplementary Fig. 1 for quantification. **b**, Egg cytosol was mock-depleted or Aurora-B-depleted and analysed by western blotting. **c**, Sperm chromatin was incubated with membranes and mock- or Aurora-B-depleted cytosol shown in **b**, in combinations with the dominant-negative p97ΔD2 mutant (8 μM hexamer) or an inhibitory anti-Ufd1 antibody (0.36 μg μl^{-1}). Note that Aurora B depletion rendered nuclear envelope formation insensitive to p97 and Ufd1 inhibition. Scale bar, 5 μm. **d**, Experiments as in **c** were quantified. Closed nuclear envelopes were scored visually on more than 50 particles ($n = 3$, bars show s.e.m., $*P < 0.05$). **e**, Nuclear envelope formation assays were performed with mock- or p97-depleted extract in combination with hesperadin (1 μM) or its solvent alone ($n = 4$, bars show s.e.m., $*P < 0.05$, $***P < 0.001$).

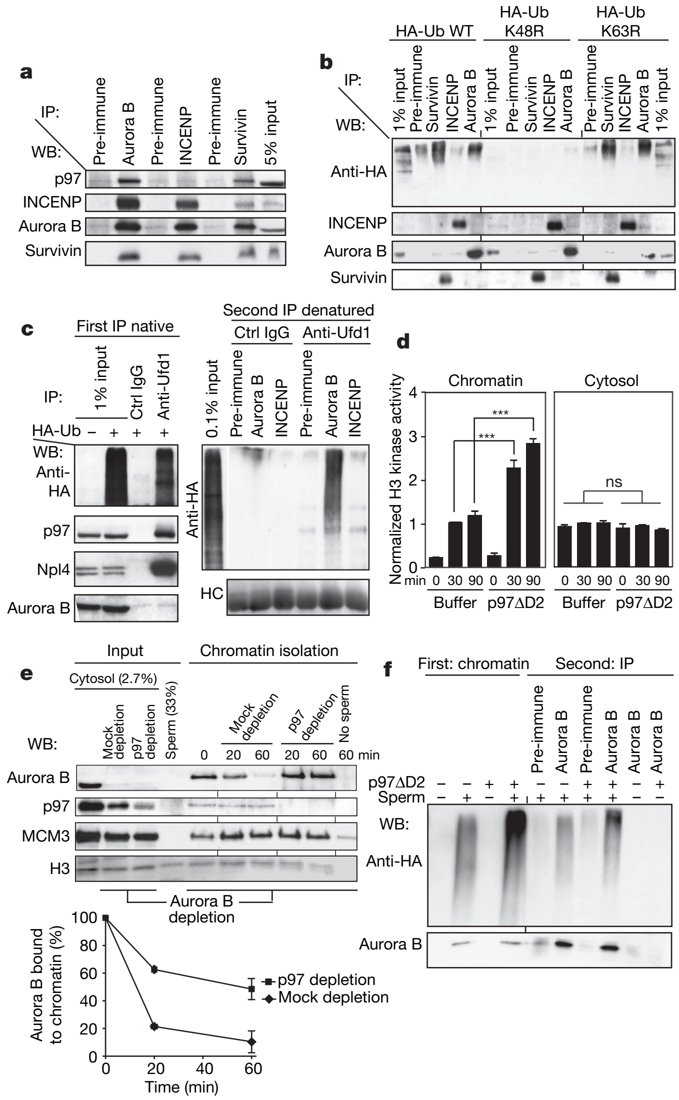


Figure 2 | p97 binds ubiquitylated Aurora B and extracts it from chromatin. **a**, Aurora B, INCENP and survivin were immunoprecipitated (IP) from *Xenopus* egg interphase cytosol under native conditions and the isolates analysed by western blotting (WB). **b**, Egg cytosol was incubated for 30 min with 20 μM HA-tagged WT ubiquitin or the chain extension mutants K48R or K63R, and denatured in 1% SDS to break protein complexes. After renaturation, Aurora B, survivin and INCENP were immunoprecipitated and analysed by western blotting. **c**, Egg cytosol was incubated with 95 μM HA-ubiquitin and subjected to anti-Ufd1 or control immunoprecipitation. One half was directly analysed by western blotting (first IP native). The other half was denatured, split into three and subjected to immunoprecipitation with indicated antibodies (second IP denatured). Note that only ubiquitylated Aurora B co-isolated with p97$^{\mathrm{Ufd1-Npl4}}$. **d**, Sperm chromatin was incubated with cytosol in the absence or presence of p97ΔD2 (8 μM) followed by re-isolation of the chromatin. Histone H3 kinase activity in cytosol and on chromatin was determined and normalized to the activity at 30 min on chromatin or in cytosol, respectively ($n = 3$, bars show s.e.m., $***P < 0.001$; ns, not significant). **e**, Sperm chromatin was incubated with complete cytosol to load it with Aurora B and isolated through a sucrose cushion (0 min). It was then incubated with Aurora-B-depleted cytosol that was either mock- or p97-depleted, and then re-isolated at indicated times for analysis by western blotting with antibodies or Ponceau stain (H3). MCM3 was probed to control chromatin recovery. Lower panel, quantification of western blot signal ($n = 3$, bars show s.e.m.). **f**, Chromatin was incubated in egg cytosol with HA-ubiquitin in the absence or presence of p97ΔD2, and then re-isolated; 2.5% was directly analysed by western blotting (first: chromatin). From the rest, proteins were salt-eluted, subjected to Aurora B or control immunoprecipitations, and the isolates analysed as indicated (second: IP).

specifically removes and inactivates chromatin-associated Aurora B. In fact, Aurora B kinase is only fully active when clustered by recruitment to chromatin or other structures[23]. We therefore asked whether p97 inactivates Aurora B on chromatin through dissociation of the clusters. To monitor Aurora B dissociation from chromatin specifically, we first loaded chromatin with Aurora B in complete cytosol, re-isolated the chromatin and incubated it with Aurora B-depleted cytosol to prevent rebinding to chromatin from the soluble pool. Under these conditions, the level of chromatin-associated Aurora B decreased rapidly (Fig. 2e). This was not due to passive dissociation, because depletion of ATP prevented removal of Aurora B (Supplementary Fig. 4). Crucially, depletion of p97 from Aurora-B-depleted cytosol impaired Aurora B dissociation (Fig. 2e). Furthermore, addition of the monoclonal Ufd1 antibody or p97ΔD2, but not of the inactive p97ΔD2(K251A) mutant, also inhibited Aurora B dissociation (Supplementary Fig. 4). This suggests that p97$^{\mathrm{Ufd1-Npl4}}$ mediates mobilization of the kinase from chromatin. To test whether Aurora B mobilization involved ubiquitylation, we incubated chromatin with p97ΔD2. Subsequent chromatin isolation, salt-elution of associated proteins and re-isolation of Aurora B revealed that ubiquitylated Aurora B accumulated on chromatin in the presence of p97ΔD2, indicating that p97 mobilizes Aurora B after its ubiquitylation (Fig. 2f).

To ask whether p97-mediated extraction of Aurora B played a role in nucleus formation during exit from mitosis, we incubated sperm chromatin in metaphase-arrested egg extract (cytostatic factor (CSF)-extract). The chromatin condensed with clearly visible chromatids and high levels of histone H3 phosphorylation (Fig. 3a). Exit from mitosis was initiated by addition of Ca^{2+}. In controls, after a delay of about 20 min, H3 was quickly dephosphorylated and chromatin decondensed (Fig. 3a–c). Addition of hesperadin resulted in an immediate drop of H3 phosphorylation and fast decondensation without delay (Fig. 3a–c), suggesting that Aurora B remained active and maintained condensation during the first 20 min (approximately) in the control. Consistently, chromatin decondensation was also prevented by addition of active Aurora B (Fig. 3f). Inactivation of p97 with p97ΔD2, too, led to an inhibition of chromatin decondensation that was accompanied with persisting Aurora B on chromatin (Fig. 3d) and delay in H3 dephosphorylation (Fig. 3a–c), indicating that p97 activity is needed to remove Aurora B from chromatin also during exit from mitosis. This was not due to a block of mitotic exit, because Cdk1 was normally inactivated (Fig. 3e). Importantly, both decondensation and H3 dephosphorylation were restored by inactivation of Aurora B with hesperadin in the presence of p97ΔD2 (Fig. 3a–c). Nuclear envelope formation coincided with decondensation and H3 dephosphorylation, as it was blocked by p97 inhibition or Aurora B addition, and restored with hesperadin (Fig. 3f–g). These data indicate that p97 needs to extract Aurora B from chromatin to allow both chromatin decondensation and nuclear envelope formation.

Next, we asked whether p97 regulates nucleus formation *in vivo*. We studied the process in *C. elegans* embryos that expressed the signal-peptide peptidase SP12 fused to green fluorescent protein (GFP) as an ER and nuclear envelope marker[24]. In WT embryos, nuclear envelope membrane formed early around decondensing chromatin and appeared as a clear ring in interphase (Fig. 4a). Nuclear pore complexes assembled concurrently and were largely absent from the cytoplasm. Co-depletion of the two redundant homologues of p97, CDC-48.1 and CDC-48.2 (ref. 25) by injection of double-stranded RNA (dsRNA) resulted in 100% embryonic lethality ($n > 100$ embryos). Western blot analysis confirmed that both CDC-48.1 and CDC-48.2 are efficiently depleted upon RNA-mediated interference (RNAi) (Supplementary Fig. 5). Analysis of early embryos showed strong defects in reformation of nuclei. In 14 out of 33 embryos, all chromatin remained condensed and lacked nuclear envelope compared with 1 out of 13 in WT. Almost half of the embryos (15 out of 33 compared with 3 out of 13 in WT) showed an intermediate phenotype with some nuclei, but also condensed chromatin without apparent nuclear envelope. Strikingly, components of the nuclear pore complex were scattered throughout the cytoplasm (29 out of 33), most likely assembling aberrantly in the ER because of the lack of nuclear envelope formation at the end of mitosis. These data demonstrate that CDC-48/p97 is critical for the formation of the nucleus also *in vivo*.

Like in egg extracts, depletion of CDC-48 led to accumulation of Aurora B protein on chromatin in embryos (Fig. 4b and Supplementary Fig. 6a). To test whether CDC-48 also regulates Aurora B activity, we studied H3 phosphorylation in a temperature-sensitive mutant of Aurora B, *air-2(or207)* (ref. 26). In almost

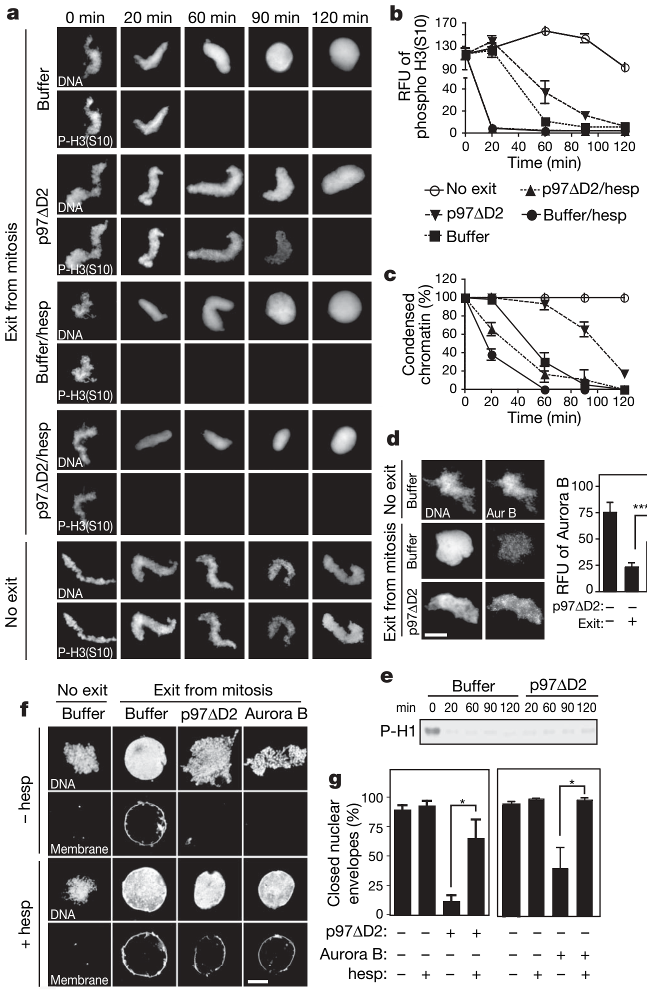


Figure 3 | p97-regulated Aurora B governs chromatin decondensation and nuclear envelope formation during exit from mitosis. **a**, Sperm chromatin was incubated for 45 min in metaphase-arrested egg extract (CSF-extract) to generate mitotic chromatin (0 min). The extract was released from mitosis by Ca^{2+} addition in the presence of buffer, p97ΔD2 (8 μM), hesperadin (1 μM) or combinations. A reaction without addition of Ca^{2+} served as negative control (no exit). Chromatin was fixed at indicated times, spun on coverslips and stained with DAPI and anti-phospho-H3(S10) antibodies. The relative fluorescence (RFU) of indirect phospho-H3 immunostaining (**b**) and the percentage of condensed chromatin particles as judged by disappearance of chromatid threads (**c**) was determined on more than 50 chromatin particles each in three independent experiments as in **a**. **d**, Indirect immunofluorescence of Aurora B on chromatin 60 min after addition of Ca^{2+} in the presence of p97ΔD2 or buffer alone. **e**, Histone H1 kinase activity during exit from mitosis in the presence or absence of p97ΔD2. **f**, Chromatin from experiments as in **a** was fixed at 90 min, stained with DAPI and $DiOC_6$, and imaged by confocal microscopy. Aurora B was added at 5 μM. Scale bar, 10 μm. **g**, Quantification of nuclear envelope formation after 120 min (Aurora B, 3.5 μM) **b**, **c**, **d**, **g** $n = 3$, bars show s.e.m., $^{*}P < 0.05$ $^{***}P < 0.001$.

none of the mutant embryos grown at semi-permissive temperature could we detect phosphorylation of H3 on mitotic chromatin (5% positive compared with 52% in WT), whereas decondensation of the chromatin was not impaired (Fig. 4c and Supplementary Fig. 6b). Depletion of CDC-48 again led to accumulation of condensed chromatin and inhibition of nuclear envelope formation, but also to an increase of embryos with phospho-H3-positive chromatin (96%). Importantly, depletion of CDC-48 in *air-2(or207)* mutants restored H3 phosphorylation (87%), whereas control RNAi directed against an unrelated gene adjacent to *cdc-48.2*, *F10B5.8*, neither interfered with nucleus formation nor restored H3 phosphorylation (9% positive, micrograph not shown). This indicates that p97 is an important negative regulator of Aurora B *in vivo*. This notion was supported by our finding that feeding *air-2(or207)* mutant worms with bacteria expressing dsRNA to either *cdc-48.1* or *cdc-48.2* suppressed embryonic lethality from 97% to 63% and 78%, respectively. The nucleus formation defects induced by depletion of CDC-48 were not significantly restored in *air-2(or207)*, which may be because of high Aurora B activity in the mutant after CDC-48 depletion (Fig. 4c). For that reason, the relevance of Aurora B regulation by CDC-48 for nucleus reformation in worm embryos cannot be assessed at this point.

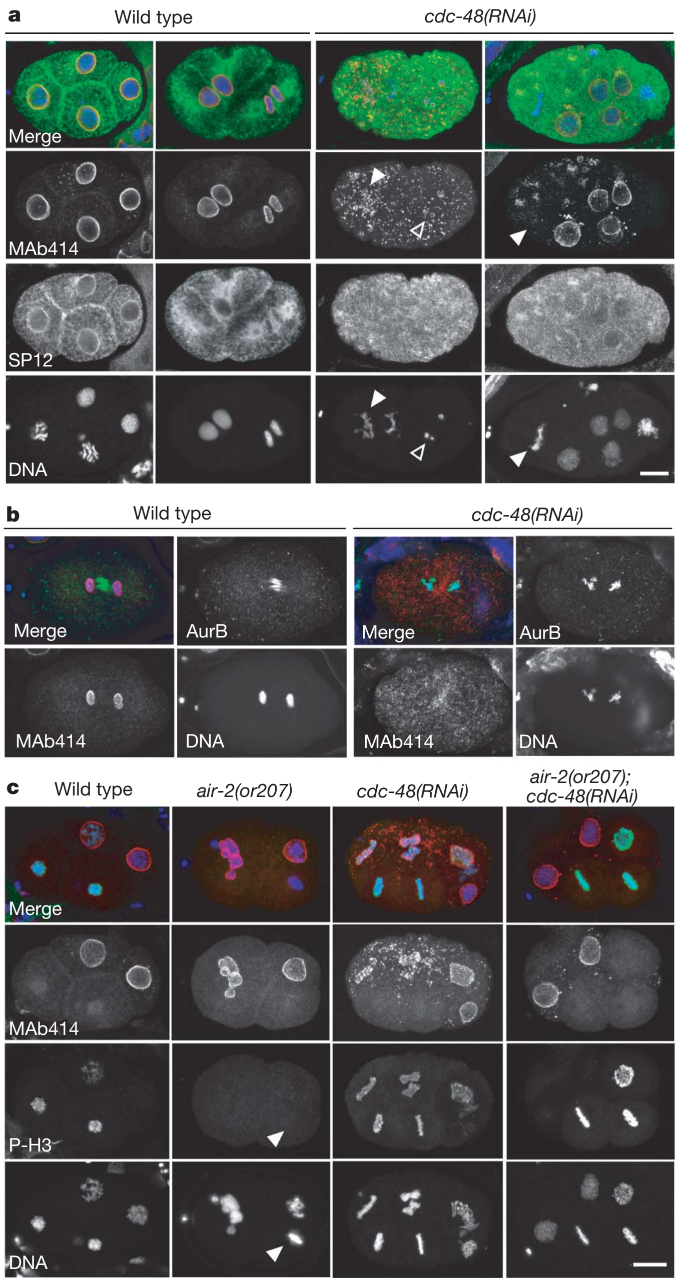


Figure 4 | CDC-48/p97 regulates nucleus formation and Aurora B on chromatin in *C. elegans*. **a**, WT and *cdc-48.1(RNAi) cdc-48.2(RNAi)* embryos expressing the ER/nuclear envelope marker SP12::GFP. Embryos were fixed and analysed by indirect immunofluorescence with the MAb414 antibody to nucleoporins and anti-GFP. DNA was visualized with DAPI. Note that chromatin frequently remained condensed without nuclear envelope or nucleoporin recruitment (filled arrowhead) or only formed small nuclei (open arrowhead). **b**, WT and *cdc-48.1(RNAi) cdc-48.2(RNAi)* embryos were fixed and stained with Aurora B/AIR-2 antibodies. See Supplementary Fig. 6a for quantification. **c**, CDC-48 depletion restores AIR-2 activity. WT or *air-2(or207)* mutant worms were injected with dsRNA to *cdc-48.1* and *cdc-48.2* and incubated at 18 °C for 24 h. Embryos were fixed and stained with anti-phospho-histone H3(S10) and with MAb414, as well as with the DNA dye DAPI. Note the phospho-H3 negative mitotic chromatin (filled arrowheads). See Supplementary Fig. 6b for quantification. Scale bar, 10 μm.

Taken together, our results provide evidence for a novel crucial pathway that regulates reformation of the nucleus after mitosis in embryonic cells. They suggest that, during mitosis, Aurora B kinase inhibits formation of the nucleus by preventing chromosome decondensation and nuclear envelope formation. It will be interesting to clarify whether this is caused by phosphorylation of known Aurora B substrates, histone H3 or subunits of condensin I[11–13,15]. During exit from mitosis, the ubiquitin-selective chaperone complex $p97^{Ufd1–Npl4}$ inactivates Aurora B on chromatin, thus allowing Aurora B substrates to be dephosphorylated and the nucleus to assemble. Previous data suggested that Ufd1 might have a positive role in Aurora B function during chromosome segregation in human somatic cells[10]. Our biochemical and genetic data, however, clearly support the notion that the $p97^{Ufd1–Npl4}$ complex functions as a negative regulator of Aurora B at least in embryonic cells.

Because p97 binds to polyubiquitylated Aurora B and lack of p97 activity leads to accumulation of ubiquitylated Aurora B on chromatin, we propose that $p97^{Ufd1–Npl4}$ inactivates Aurora B by extracting it from chromatin in a similar manner as it extracts ubiquitylated proteins from the ER membrane. This may be coupled to extraction of the Aurora B cofactor survivin, which we found is also ubiquitylated. So far, it is unclear whether extracted Aurora B is degraded or, as with substrates in other p97-mediated processes[5], recycled by de-ubiquitylation, which would be consistent with bulk levels of Aurora B being stable in embryos. In conclusion, our results identify for the first time a ubiquitylated substrate of Cdc48/p97 that is relevant for organelle biogenesis after mitosis. They also reconcile the role of Cdc48/p97 in nucleus formation with its function in other ubiquitin-dependent processes in interphase. Finally, given that Aurora B is regulated on chromatin, our data highlight the importance of localized ubiquitin-dependent regulation and how this controls a mitotic event.

METHODS SUMMARY

C. elegans. The WT strain was N2 (Bristol). The SP12::GFP-expressing strain[24] and *air-2(or207)* (ref. 26) were maintained at 22 °C and 15 °C, respectively. RNAi was performed by injecting dsRNA to *cdc-48.1* and *cdc-48.2* or *F10B5.8* into adults and embryos analysed after 24 h. Constructs used for RNAi are as previously described[27]. Images of methanol-fixed embryos represent maximum intensity projections of Leica SP1 confocal microscope sections.

Proteins and antibodies. His-p97, His-p97ΔD2 (residues 1–458), HA-ubiquitin and Aurora B (60–361)/INCENP (790–856) complex were prepared as previously described[19,22,28]. Antibodies were raised in rabbits to glutathione *S*-transferase (GST)-Aurora B, GST-survivin and an INCENP peptide (CSNRHHLAVGYGLKY) from *X. laevis*. Antibodies to p97 and Ufd1 (5E2) were described[28]. Other antibodies were from P. Jackson (MCM3), Covance (HA, MAb414), Upstate (phospho-H3-S10), Qiagen (RGS-His), Abcam (GFP), K. Oegema (AIR-2), K. Yamanaka and T. Ogura (CDC-48).

Extracts and *in vitro* reactions. *X. laevis* egg fractions were generated and nucleus formation reactions were performed as described[29,30]. Samples were fixed in formaldehyde and visualized with a Leica SP2 AOBS confocal or a Zeiss Axiovert 100TV epifluorescence microscope. Indirect immunofluorescence was quantified

with NIH ImageJ software. We depleted p97 using GST-UT6 as described[3]. Native immunoprecipitations and immunodepletions were performed in reaction buffer and washed in 0.1% Triton X-100. For denaturing immunoprecipitations, samples were first adjusted to 1% SDS and then diluted to 0.1% SDS with 1% Triton X-100 in PBS before isolation.

Chromatin isolation. Reactions were stopped by dilution and chromatin sedimented through a 1.5 M sucrose cushion at 10,000*g*. For kinase assays, pellets were washed and incubated in 5 μl kinase buffer for 15 min at room temperature, and reactions separated by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) and quantified with a phosphorimager. Radiolabelled RCC1 was added for normalization of chromatin amounts[16].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 October; accepted 17 October 2007.

1. Burke, B. & Ellenberg, J. Remodelling the walls of the nucleus. *Nature Rev. Mol. Cell Biol.* **3,** 487–497 (2002).
2. Hetzer, M. W., Walther, T. C. & Mattaj, I. W. Pushing the envelope: structure, function, and dynamics of the nuclear periphery. *Annu. Rev. Cell Dev. Biol.* **21,** 347–380 (2005).
3. Hetzer, M. *et al.* Distinct AAA-ATPase p97 complexes function in discrete steps of nuclear assembly. *Nature Cell Biol.* **3,** 1086–1091 (2001).
4. Ye, Y. Diverse functions with a common regulator: ubiquitin takes command of an AAA ATPase. *J. Struct. Biol.* **156,** 29–40 (2006).
5. Jentsch, S. & Rumpf, S. Cdc48 (p97): a ''molecular gearbox'' in the ubiquitin pathway? *Trends Biochem. Sci.* **32,** 6–11 (2007).
6. Kondo, H. *et al.* p47 is a cofactor for p97-mediated membrane fusion. *Nature* **388,** 75–78 (1997).
7. Burke, B. The nuclear envelope: filling in gaps. *Nature Cell Biol.* **3,** E273–E274 (2001).
8. Meusser, B., Hirsch, C., Jarosch, E. & Sommer, T. ERAD: the long road to destruction. *Nature Cell Biol.* **7,** 766–772 (2005).
9. Vagnarelli, P. & Earnshaw, W. C. Chromosomal passengers: the four-dimensional regulation of mitotic events. *Chromosoma* **113,** 211–222 (2004).
10. Vong, Q. P., Cao, K., Li, H. Y., Iglesias, P. A. & Zheng, Y. Chromosome alignment and segregation regulated by ubiquitination of survivin. *Science* **310,** 1499–1504 (2005).
11. Hagstrom, K. A., Holmes, V. F., Cozzarelli, N. R. & Meyer, B. J. C. *elegans* condensin promotes mitotic chromosome architecture, centromere organization, and sister chromatid segregation during mitosis and meiosis. *Genes Dev.* **16,** 729–742 (2002).
12. Kaitna, S., Pasierbek, P., Jantsch, M., Loidl, J. & Glotzer, M. The aurora B kinase AIR-2 regulates kinetochores during mitosis and is required for separation of homologous Chromosomes during meiosis. *Curr. Biol.* **12,** 798–812 (2002).
13. Lipp, J. J., Hirota, T., Poser, I. & Peters, J. M. Aurora B controls the association of condensin I but not condensin II with mitotic chromosomes. *J. Cell Sci.* **120,** 1245–1255 (2007).
14. Takemoto, A. *et al.* Analysis of the role of Aurora B on the chromosomal targeting of condensin I. *Nucleic Acids Res.* **35,** 2403–2412 (2007).
15. Hsu, J. Y. *et al.* Mitotic phosphorylation of histone H3 is governed by Ipl1/aurora kinase and Glc7/PP1 phosphatase in budding yeast and nematodes. *Cell* **102,** 279–291 (2000).
16. Fischle, W. *et al.* Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature* **438,** 1116–1122 (2005).
17. Hirota, T., Lipp, J. J., Toh, B. H. & Peters, J. M. Histone H3 serine 10 phosphorylation by Aurora B causes HP1 dissociation from heterochromatin. *Nature* **438,** 1176–1180 (2005).
18. Kourmouli, N. *et al.* Dynamic associations of heterochromatin protein 1 with the nuclear envelope. *EMBO J.* **19,** 6558–6568 (2000).
19. Sessa, F. *et al.* Mechanism of Aurora B activation by INCENP and inhibition by hesperadin. *Mol. Cell* **18,** 379–391 (2005).
20. Hauf, S. *et al.* The small molecule Hesperadin reveals a role for Aurora B in correcting kinetochore-microtubule attachment and in maintaining the spindle assembly checkpoint. *J. Cell Biol.* **161,** 281–294 (2003).
21. Sugiyama, K. *et al.* Aurora-B associated protein phosphatases as negative regulators of kinase activation. *Oncogene* **21,** 3103–3111 (2002).
22. Ye, Y., Meyer, H. H. & Rapoport, T. A. Function of the p97–Ufd1–Npl4 complex in retrotranslocation from the ER to the cytosol: dual recognition of nonubiquitinated polypeptide segments and polyubiquitin chains. *J. Cell Biol.* **162,** 71–84 (2003).
23. Kelly, A. E. *et al.* Chromosomal enrichment and activation of the aurora B pathway are coupled to spatially regulate spindle assembly. *Dev. Cell* **12,** 31–43 (2007).
24. Poteryaev, D., Squirrell, J. M., Campbell, J. M., White, J. G. & Spang, A. Involvement of the actin cytoskeleton and homotypic membrane fusion in ER dynamics in *Caenorhabditis elegans. Mol. Biol. Cell* **16,** 2139–2153 (2005).
25. Mouysset, J., Kahler, C. & Hoppe, T. A conserved role of *Caenorhabditis elegans* CDC-48 in ER-associated protein degradation. *J. Struct. Biol.* **156,** 41–49 (2006).
26. Severson, A. F., Hamill, D. R., Carter, J. C., Schumacher, J. & Bowerman, B. The aurora-related kinase AIR-2 recruits ZEN-4/CeMKLP1 to the mitotic spindle at metaphase and is required for cytokinesis. *Curr. Biol.* **10,** 1162–1171 (2000).
27. Kamath, R. S. *et al.* Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421,** 231–237 (2003).
28. Meyer, H. H., Shorter, J. G., Seemann, J., Pappin, D. & Warren, G. A complex of mammalian Ufd1 and Npl4 links the AAA-ATPase, p97, to ubiquitin and nuclear transport pathways. *EMBO J.* **19,** 2181–2192 (2000).
29. Murray, A. W. Cell cycle extracts. *Methods Cell Biol.* **36,** 581–605 (1991).
30. Lohka, M. J. Analysis of nuclear envelope assembly using extracts of *Xenopus* eggs. *Methods Cell Biol.* **53,** 367–395 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Y. Zheng for sharing data before publication and for the GST-survivin construct, A. Musacchio for the Aurora B/INCENP construct, K. Oegema for AIR-2 antibodies, P. Jackson for MCM3 antibodies, K. Yamanaka and T. Ogura for CDC-48 antibodies, N. Kraut (Boehringer Ingelheim, Vienna) for hesperadin, and C. Brasseur and C. Zbinden for technical help. This work was supported by grants of the Swiss National Fund, the ETH Zurich, the Bonizzi-Theler Stiftung, the Roche Research Foundations (to K.R.), the Novartis Foundation and the DAAD (to O.P.). R.B., F.M.S. and T.B. were on the Molecular Life Science PhD Program Zurich.

Author Contributions M.G. and F.M.S. performed the *C. elegans* experiments. H.H.M. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to H.M. (hemmo.meyer@bc.biol.ethz.ch).

METHODS

C. elegans. The wild-type strain was N2 (Bristol). The SP12::GFP-expressing strain[24] and *air-2(or207)* (ref. 26) were maintained at 22 °C and 15 °C, respectively. Constructs used for RNAi were as previously described[27]. Briefly, the genomic regions were amplified using the following oligonucleotides: CDC-48.1 (C06A1.1), 5′-GTGAATGCGCAGAATCTCAA-3′, 5′-ATCCTAACACC-ACGCAAAGG-3′; CDC-48.2 (C41C4.8), 5′-CGTTTAACCATGGCCTCAGT-3′, 5′-GACGAAGGGCTCCATCAATA-3′; F10B5.8, 5′-GATTACCGAAAA-GTTCAATGTG-3′, 5′-CAGCATGAGCACTAAACGAC-3′.

F10B5.8 PCR products were cloned into pDEST-L4440 and confirmed by sequencing. dsRNA was produced *in vitro* (Ribomax, Promega). For microscopic analysis of embryos, adult hermaphrodites were injected with dsRNA to *cdc-48.1* and *cdc-48.2* or *F10B5.8* and maintained at 22 °C or 18 °C as indicated, for 24 h. Embryos were methanol-fixed and stained according to standard procedures. Images were acquired using a Leica SP1 confocal microscope and presented as maximum intensity projections. For western blot analysis, L1 larvae were fed for 24–28 h on NGM plates with 3 mM IPTG at 25 °C. Adult worms were collected by washing the plates with M9 buffer, washed twice in M9 buffer and boiled for 5 min in Sample buffer.

Proteins and antibodies. Recombinant His-p97, His-p97ΔD2 (residues 1–458), His-p97ΔD2KA (residues 1–458, with a lysine-to-alanine exchange at residue 251) were generated in bacteria as described[22] except that p97ΔD2 variants were further purified on a MonoQ column (Amersham). HA-ubiquitin and Aurora B (60-361)/INCENP (790-856) complex were prepared as described elsewhere[22,31]. Antibodies were raised in rabbits to *X. laevis* GST-Aurora B, GST-survivin and an INCENP peptide (CSNRHHLAVGYGLKY). Antibodies to p97 (HME5) and Ufd1 (5E2) were described previously[28]. Other antibodies were from: P. Jackson (MCM3), Covance (HA, MAb414), Upstate (phospho-histone H3(S10)), Qiagen (RGS-His), Abcam (GFP), K. Oegema (AIR-2)[32], K. Yamanaka and T. Ogura (CDC-48)[33].

Extracts and *in vitro* reactions. *X. laevis* egg fractions were generated and reactions of nuclear envelope formation were performed essentially as described elsewhere[29,30]. Briefly, eggs were lysed by centrifiguation at 37,500*g* in ELB buffer (10 mM HEPES pH 7.7, 250 mM sucrose, 50 mM KCl, 2.5 mM $MgCl_2$, 1 mM DTT) and the intermediate fraction recovered as interphase egg extract. CSF-extracts were generated by lysis in XB buffer (10 mM HEPES pH 7.7, 50 mM sucrose, 100 mM KCl, 0.1 mM $CaCl_2$, 1 mM $MgCl_2$) containing 5 mM EGTA. Where indicated, extracts were further separated into cytosol and light membranes by centrifugation for 70 min at 250,000*g*.

Typically, 20-μl reactions containing indicated egg fractions, de-membranated sperm chromatin heads (approximately 250 per microlitre) and ATP-regenerating system (2 mM ATP, 10 mM creatine phosphate, 20 μg ml^{-1} creatine kinase, 0.1 mM GTP) were incubated for indicated times. CSF-extract was activated by addition of 0.5 mM $CaCl_2$.

For morphological analysis of nucleus formation, samples were formaldehyde-fixed in phosphate-buffered saline, stained with DAPI and $DiOC_6$, and visualized with a Leica SP2 AOBS confocal microscope. For quantification, at least 50 nuclei were scored visually using a Zeiss Axiovert 100TV epifluorescence microscope. For indirect immunofluorescence, chromatin was formaldehyde-fixed, spun on coverslips, and decorated with antibodies in 0.1% Triton X-100. Epifluorescence pictures of at least 50 chromatin particles per data point were taken at identical settings, and mean intensities quantified with NIH ImageJ software.

Ubiquitylation reactions were done in egg cytosol, ATP-regenerating system and HA-ubiquitin at indicated concentrations for 30 min at 19 °C. The reaction was stopped by addition of fresh 5 mM N-ethylmaleimide.

Depletions and immunoprecipitations. Aurora B was immuno-depleted from cytosol by using rabbit antibodies and the corresponding preimmune antibodies as control. Depletion of p97 was done as described before[3]. Native immunoprecipitations were performed in XB buffer and precipitates washed in WB (50 mM HEPES pH 7.4, 120 mM NaCl, 1 mM $MgCl_2$, 1 mM EGTA and 0.1% Triton X-100). For denaturing immunoprecipitations, samples were first adjusted to 1% SDS and then diluted to 0.1% SDS in PBS containing 1% BSA and 1% Triton X-100 before subsequent immunoprecipitations.

Chromatin isolation assays. For Aurora B mobilization assays, sperm chromatin was preincubated in cytosol at 19 °C for 20 min. The reaction was diluted in ELB, and chromatin sedimented through a 1.5 M sucrose cushion at 10,000*g* for 10 min in a swing-out rotor. The chromatin was recovered and re-incubated with manipulated cytosol as indicated. Subsequently, it was re-isolated by sedimentation for western blotting analysis. Detection was done using chemiluminescence that was quantified directly with an AlphaInnotech Fluorchem 8900 imager. For ubiquitylation assays, chromatin was incubated in cytosol containing HA-ubiquitin for 30 min and isolated. Chromatin-associated proteins were eluted in high-salt buffer (600 mM NaCl, 10 mM HEPES, 0.5% NP-40, pH 7.4) for 10 min on ice. The eluate was diluted tenfold in buffer (110 mM KCl, 10 mM HEPES, 0.5% NP-40, 1% BSA pH 7.4) and subjected to immunoprecipitations.

Kinase assays. RCC1 (DNA provided by U. Kutay, ETH Zurich) was translated *in vitro* and labelled with ^{35}S in reticulocyte lysate (Promega). Egg cytosol was incubated with ATP-regenerating system, de-membranated sperm nuclei (600 per microlitre) and 1 μl RCC1 in 60 μl at 19 °C for indicated times. Reactions were stopped by dilution in 1 ml buffer (50 mM KCl, 10 mM HEPES, 1 mM DTT, 2.5 mM $MgCl_2$, 250 mM sucrose, 80 mM β-glycerophosphate, pH 7.5), and chromatin isolated as described above. Pellets were washed and incubated in 5 μl kinase buffer (80 mM β-glyerophosphate, 15 mM $MgCl_2$, 20 mM EGTA, 1 mM Na_3VO_4, 5 mM NaF and 0.3 mM ATP, 100 μM γ-^{32}P-ATP and 3 μg calf thymus H3 or H1) for 15 min at room temperature, and reactions separated by SDS-gels and quantified with a phosphorimager (Molecular Dynamics). Histone H3 phosphorylation activity was normalized to the RCC1 signal in each lane to control for chromatin recovery.

31. Wang, Y., Satoh, A., Warren, G. & Meyer, H. H. VCIP135 acts as a deubiquitinating enzyme during p97-p47-mediated reassembly of mitotic Golgi fragments. *J. Cell Biol.* **164**, 973–978 (2004).
32. Audhya, A. *et al.* A complex containing the Sm protein CAR-1 and the RNA helicase CGH-1 is required for embryonic cytokinesis in *Caenorhabditis elegans*. *J. Cell Biol.* **171**, 267–279 (2005).
33. Yamauchi, S., Yamanaka, K. & Ogura, T. Comparative analysis of expression of two p97 homologues in *Caenorhabditis elegans*. *Biochem. Biophys. Res. Commun.* **345**, 746–753 (2006).

Vol 450 | 20/27 December 2007 | doi:10.1038/nature06389

LETTERS

Visualizing spatially correlated dynamics that directs RNA conformational transitions

Qi Zhang[1], Andrew C. Stelzer[1], Charles K. Fisher[1] & Hashim M. Al-Hashimi[1]

RNAs fold into three-dimensional (3D) structures that subsequently undergo large, functionally important, conformational transitions in response to a variety of cellular signals[1–3]. RNA structures are believed to encode spatially tuned flexibility that can direct transitions along specific conformational pathways[4,5]. However, this hypothesis has proved difficult to examine directly because atomic movements in complex biomolecules cannot be visualized in 3D by using current experimental methods. Here we report the successful implementation of a strategy using NMR that has allowed us to visualize, with complete 3D rotational sensitivity, the dynamics between two RNA helices that are linked by a functionally important trinucleotide bulge over timescales extending up to milliseconds. The key to our approach is to anchor NMR frames of reference onto each helix and thereby directly measure their dynamics, one relative to the other, using 'relativistic' sets of residual dipolar couplings (RDCs)[6,7]. Using this approach, we uncovered super-large amplitude helix motions that trace out a surprisingly structured and spatially correlated 3D dynamic trajectory. The two helices twist around their individual axes by approximately 53° and 110° in a highly correlated manner ($R = 0.97$) while simultaneously ($R = 0.81$–0.92) bending by about 94°. Remarkably, the 3D dynamic trajectory is dotted at various positions by seven distinct ligand-bound conformations of the RNA. Thus even partly unstructured RNAs can undergo structured dynamics that directs ligand-induced transitions along specific predefined conformational pathways.

Although advances in solution NMR methods are providing rare insights into the atomic details of internal motions in biomolecules[8–11], the 3D visualization of atomic movements remains hampered by the fact that common NMR measurements do not probe the dynamics of bond vectors relative to one another, but rather, relative to the external magnetic field. Reliance on a single external frame of reference reduces the possibilities for overcoming fundamental spatial sensitivity limitations associated with axially symmetric interactions, makes it impossible to establish spatial correlations between motions occurring at different sites, and renders characterization of internal motions only possible if their spectroscopic contributions can be disentangled from those due to the much larger overall brownian motions[12]. The last proves intractable when domains in multi-domain systems move collectively and lead to correlated changes in the overall motions, as is generally encountered when A-form helices move in RNA[13].

We transformed the basic NMR experiment by anchoring frames of reference onto individual RNA helices and thereby measuring their dynamics directly as motions of one helix relative to the other by using multiple sets of RDCs. The frames were anchored using a generally applicable method which involves elongating helices[13] so that they dominate overall alignment of the elongated RNA in ordering media, with the elongated axis being on average oriented parallel to the magnetic field (Fig. 1a). This effectively disentangles helix motions from overall alignment and renders RDCs dependent on the angle (Ω) between bond vectors and the internal elongated axis, and not a detached external magnetic field. By anchoring the frame of reference onto different helices, the same helix motions can be measured from different helix-centred perspectives, opening a new avenue for extending the achievable spatial sensitivity with which motions can be characterized. For example, although RDCs probe inter-helical bending (β) and twisting (α) motions of the short helix, they are insensitive to twisting motions (γ) around the axially

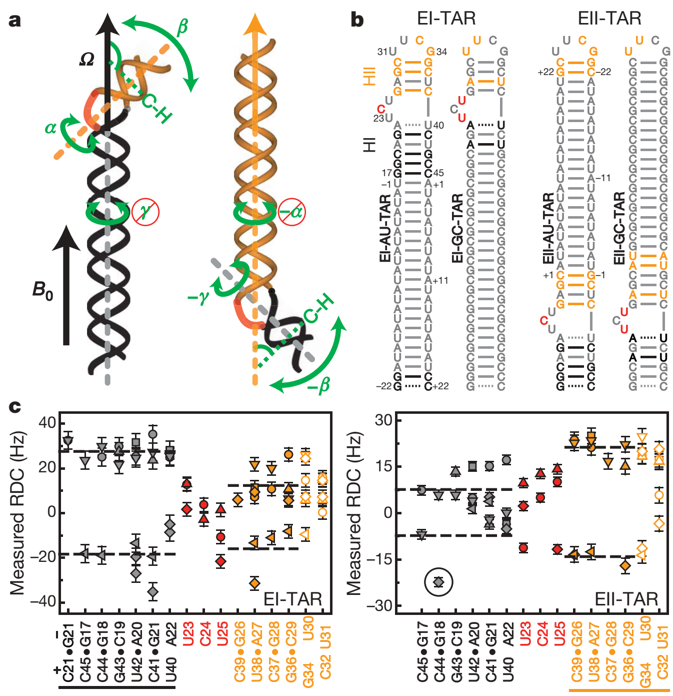


Figure 1 | Measurement of RNA helix motions in 3D using helix-anchored frames and RDCs. **a**, NMR reference frames are anchored by elongating helices so that the RNA aligns with the elongated axis, being on average oriented parallel to the magnetic field (B_0). **b**, Independent elongation of HI and HII in TAR using a strategy that renders elongation residues NMR invisible[13]. $^{13}C/^{15}N$ labelled and unlabelled nucleotides are shown in colour and grey, respectively. **c**, The C1′H1′ (diamond), C2H2 (square), C5H5 (circle), C6H6 (triangle up), C8H8 (triangle down) and N1/3H1/3 (triangle left) RDCs as a function of the TAR secondary structure. Elongated helices are underlined. The horizontal dashed lines correspond to the average positive or negative RDC value in each helix. Error bars represent experimental uncertainty (one s.d.) estimated from duplicate measurements and analysis of signal-to-noise and line widths.

[1]Department of Chemistry and Biophysics, University of Michigan, 930 North University Avenue, Ann Arbor, Michigan 48109-1055, USA.

symmetric elongated helix (Fig. 1a). This vanishing sensitivity can be resurrected, and helix motions thereby measured with complete 3D rotational sensitivity, by inverting which helix is elongated and measuring RDCs that probe the angles β and γ but not α (Fig. 1a). By measuring motions in this relativistic manner, the helix-anchored frames also allow spatial correlations between motions of two or more helices to be directly established. A similar strategy can be implemented without the need for helix elongation through analytical treatment of couplings between helix motions and overall magnetic alignment of the non-elongated RNA[14].

The transactivation response element (TAR) RNA (Fig. 1b) from the human immunodeficiency virus type-1 (HIV-1) is a major drug target and a paradigm for understanding ligand-induced RNA conformational transitions. Numerous studies have shown that HIV-1 TAR undergoes dramatic conformational rearrangements involving large rigid-body movements of its two helical domains (averaging 42° and as large as 71°) that allow it to bind diverse targets in and around the bulge, including peptide derivatives of its cognate protein Tat[15–17], divalent ions[18] and five, different, small molecules that inhibit the TAR–Tat interaction[19–22]. We recently reported evidence based on elongation of helix I and NMR spin relaxation measurements[13] for diffusion-limited collective helix motions in TAR occurring at timescales of less than 19 ns. Here, we apply our double-elongation RDC approach to visualize the TAR helix motions in 3D over timescales extending up to milliseconds.

Each of the two helices in TAR was independently elongated (EI-TAR and EII-TAR) by using a generally applicable 'NMR invisible' strategy[13] which entailed preparation of four E-TAR constructs (Fig. 1b). RDCs were measured by using Pf1 phage (approximately 6–8 mg ml^{-1}) as an ordering medium[23]. Comparison of NMR chemical shifts indicated that domain elongation and the introduction of Pf1 phage do not significantly perturb the inter-helical conformational dynamics and functional properties of TAR (Supplementary Fig. 1). Inspection of the raw RDCs measured in E-TAR immediately revealed evidence for collective helix motions (Fig. 1c). In both EI-TAR and EII-TAR, the RDCs measured for differently oriented bond vectors in the short helices were significantly smaller (on average by about 55% and about 40%, respectively) than corresponding values measured in the elongated helices (the circled RDC in EII-TAR is consistent with this trend; the large value arises because the bond vector is oriented uniquely parallel to long axis of helix II (HII)) (Fig. 1c). This indicated that the short helices align to a lesser extent than their elongated counterparts because they undergo collective helix motions at sub-millisecond timescales. The small RDCs observed at the bulge are also consistent with previous findings showing that the inter-helical linker is highly flexible[13]. As expected, the collective helix motional amplitudes and thus the observed gap in helix RDCs were significantly diminished after complexation with the ligand argininamide (ARG), which has previously been shown to stabilize the relative alignment of the two TAR helices[13] and upon shortening the bulge linker through deletion of residue C24 (Supplementary Fig. 2).

Domain-elongation allowed us to analyse the RDCs by using a model-free order-tensor analysis (Supplementary Information)[24,25] from which we obtained geometrical information about the inter-helical motional trajectory. The RDCs and previously validated idealized A-form helix geometry[26] were used to determine five order-tensor elements[27] describing ordering of each helix relative to the external magnetic field, or in the case of elongated RNA, relative to the internal elongated helix axis. An excellent RDC fit was obtained in each case, confirming that the helices also adopt the idealized A-form geometry in the elongated TAR context (Fig. 2a). The order tensors obtained for the elongated helices show that they align in the expected axially symmetric manner ($\eta \approx 0$, ranges between 0 and 1 for minimum and maximum motional asymmetry), with their elongated axis (z direction) oriented on average nearly parallel to the magnetic field (S_{zz}, deviations fewer than six degrees) (Fig. 2b).

The order tensors obtained for the short helices, on the other hand, provided insights into their orientational dynamics relative to the elongated helices (Fig. 2c). According to the short HII order tensor, the long axis (S_{zz}) of HI is on average inclined at an inter-helical bend angle of approximately 25° (Fig. 2c). In contrast, the same bend angle is approximately 54° according to the short HI order tensor (Fig. 2c). These contrasting views of the same inter-helical angle suggest the presence of twisting motions that alter how each helix views the other's average orientation. The very high degree of attenuation in the level of order (ϑ)[25] observed for the two short helices compared with their elongated counterparts ($\vartheta_{int} = \vartheta_{short}/\vartheta_{elongated} \approx 0.45 \pm 0.05$ and 0.47 ± 0.02 for EI-TAR and EII-TAR, respectively, and ranges between 0 and 1 for maximum and minimum motions) confirmed the presence of super-large amplitude helix motions that significantly exceed the amplitudes observed previously by spin relaxation ($\sqrt{S_s^2} = \vartheta_{int} \approx 0.86 \pm 0.02$)[13]. Similar motional amplitudes were observed when anchoring the frames of reference in non-elongated TAR by using motional couplings[14].

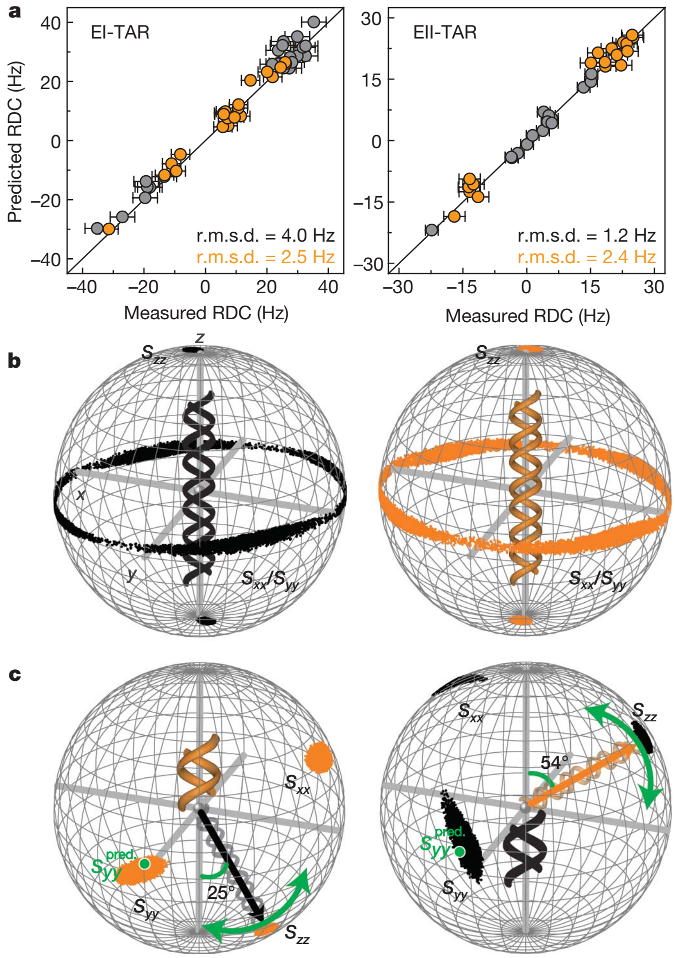


Figure 2 | Helix motions from model-free order-tensor analysis of RDCs. **a**, Comparison of RDCs measured in HI (black) and HII (orange) in EI-TAR and EII-TAR with values back-predicted by using the best-fit order tensor and an idealized A-form helix geometry[26]. Shown are the root mean squared deviations (r.m.s.d.) between measured and predicted RDCs. Error bars indicate the s.d. **b**, **c**, Globes showing the orientation of the order-tensor frames determined for the (**b**) elongated and (**c**) short helices in EI-TAR and EII-TAR. Solutions are depicted relative to a molecular frame in which HI and HII are coaxial with the helix axis oriented along the z direction. Green circles are the S_{yy} direction predicted for an inter-helix motional model involving inter-conversion among seven equally populated ligand-bound TAR conformations.

Given the broader timescale sensitivity of RDCs (less than 0.1 ms) compared with spin relaxation (less than 19 ns) and that qualitative analysis of the TAR resonance intensities provides no evidence for significant exchange broadening in and around the bulge[13], the motions likely represent large-scale diffusive dynamics occurring at the nanosecond to microsecond timescale and not microsecond to millisecond transitions between discrete sub-states. As expected, considerably smaller inter-helix motional amplitudes were observed in the EI-TAR+ARG complex ($\vartheta_{\mathrm{int}} = 1.09 \pm 0.08$ and $S_s = 1$) or after deletion of bulge residue C24 ($\vartheta_{\mathrm{int}} = 0.77 \pm 0.04$ and $S_s = 0.94 \pm 0.02$). Surprisingly, these very large amplitude helix motions observed in E-TAR are not spatially random as both short helices report a high degree of motional asymmetry ($\eta \approx 0.44$, Supplementary Table 2) and a preference for inter-helical bending through rotations around a principal direction (S_{yy}) that is nearly orthogonal to both helical axes (Fig. 2c). Remarkably, a virtually identical axis of motional asymmetry (Fig. 2c, in green circles) relates the orientation of TAR helices in seven distinct ligand-bound conformations[16–22], suggesting fundamental spatial similarities between how the helices move in free TAR and how they move when adapting to bind different targets. In contrast, the motions observed following deletion of bulge residue C24 in EI-TAR appear to be more spatially isotropic ($\eta \approx 0.07$) suggesting that the length of the bulge codes for both the amplitude and directionality of the helix motions.

To visualize the spatially non-random inter-helical motions, we performed a search over ensembles[28] with up to three ($N = 3$) equally populated inter-helical conformers that can reproduce the nine independent parameters afforded by the RDCs (Supplementary Information). The conformers within an ensemble define key loci along the trajectory that serve to capture its essential 3D spatial features even if the trajectory was to involve many more conformations. Loose steric constraints were also implemented in the conformational search (Supplementary Information). In the case of TAR, both the $N = 1$ and $N = 2$ searches yielded a very poor RDC fit, confirming the existence of an inter-helix motional trajectory that is more complex than a simple two-state jump. In contrast, a very good fit was obtained for $N = 3$, with insignificant improvements obtained with $N = 4$ (Fig. 3a).

Cluster analysis of the $N = 3$ ensemble solutions revealed two RDC-degenerate solutions (A and B) that differ primarily in whether helices in the three conformers twist in a clockwise (A) or anticlockwise (B) manner (Fig. 3b). A remarkable feature common to both solutions is that the three conformers fall nearly along a straight line in the 3D inter-helix Euler space defining twisting around each helix (α and $-\gamma$) and inter-helical bending (β) (Fig. 3b). This 3D linear trajectory cannot be traced out by a 'single axis rotation', explaining why a two-state jump does not yield an acceptable RDC fit. It suggests that the two helices twist and bend in a highly correlated manner. Although we cannot exclude the possibility that ensembles A and B co-exist with equal populations as this yields an equally good E-TAR RDC fit, cross-validation against an independent set of phage-induced RDCs measured in non-elongated TAR[29] strongly argues in favour of ensemble A over ensemble B or combination of A and B (Supplementary Information). The A ensemble, which also yields good agreement with magnetic-field-induced RDCs measured in non-elongated TAR[14], gives rise to a motional trajectory in which HI and HII twist by approximately 53° and approximately 110°, respectively, in a highly synchronized manner ($R = 0.97$) while simultaneously ($R = 0.81$–0.92) bending by about 94° (Fig. 3c, Supplementary Movie 1).

Visualization of the TAR inter-helical trajectory allowed us to examine directly its potential role in directing ligand-induced structural transitions. Specifically, it allowed us to test directly the possibility that ligands bind to existing TAR conformations by 'tertiary capture'[4,5]. Remarkably, we find that the seven bound TAR conformations fall along various positions of the A dynamical trajectory (Fig. 4a). The bound conformers also trace out a similar linear trajectory in the 3D inter-helix Euler space ($R = 0.70$–0.82), confirming that correlated dynamics is an intrinsic property of the TAR structure (Fig. 4a). The TAR dynamical envelope encapsulates nearly all of the ligand-bound conformations (Fig. 4b, Supplementary Movie 2), indicating that ligands can induce the TAR structural transitions by capturing existing conformers along various positions of the free RNA dynamical trajectory. Thus a highly flexible RNA can be spatially tuned to undergo structured motions that direct functional transitions along specific pathways. The presented NMR strategy provides a general approach to characterize collective movements of helices and other locally well defined RNA sub-fragments across a variety of functionally important junctions with 3D spatial resolution.

METHODS SUMMARY

E-TAR samples (approximately 0.6–1.0 mM) were prepared by *in vitro* transcription as described previously[13]. All experiments were conducted in NMR buffer (15 mM sodium phosphate, 25 mM sodium chloride, 0.1 mM EDTA,

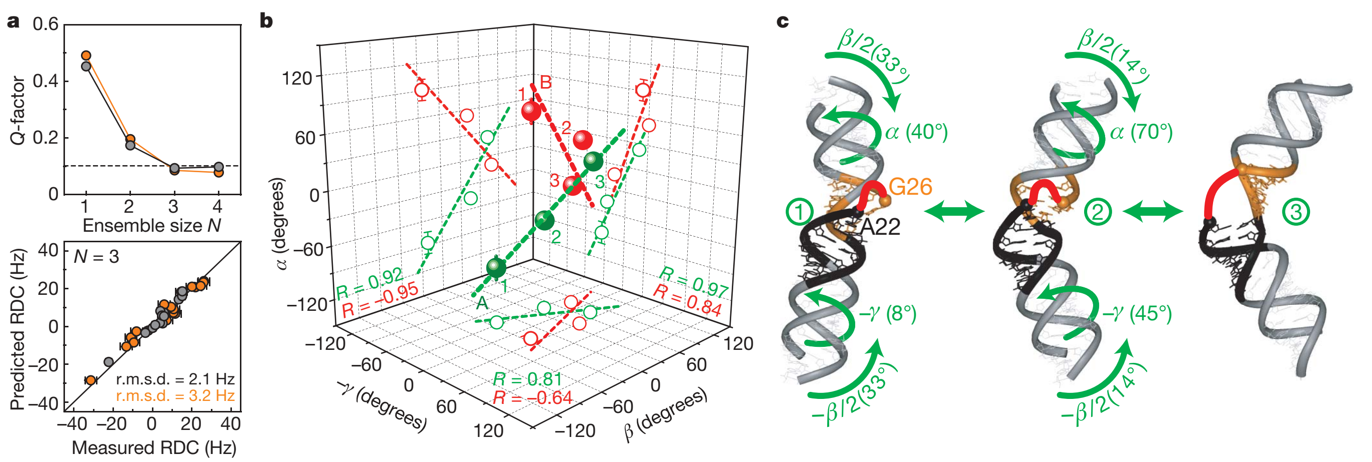


Figure 3 | Visualizing TAR inter-helical motions in 3D. **a**, Quality factor (Q)[31] for HI (black) and HII (orange) as a function of the inter-helical ensemble size N (top panel). The horizontal dashed line corresponds to the expected value of Q given experimental uncertainty. Comparison of the RDCs measured in short HI (black) and HII (orange), and values back-predicted by using the best-fit ($N = 3$) dynamical ensemble A (bottom panel). **b**, The best-fit TAR ensembles A (green) and B (red) in which each conformer (1, 2 and 3) is specified by three Euler angles defining twisting around HI ($-\gamma$), HII (α) and inter-helical bending (β). The 3D best-fit line is shown with its 2D projections along each plane and the associated correlation coefficient (R). **c**, The three TAR conformers in ensemble A. Helices are elongated for clarity. Error bars indicate the s.d.

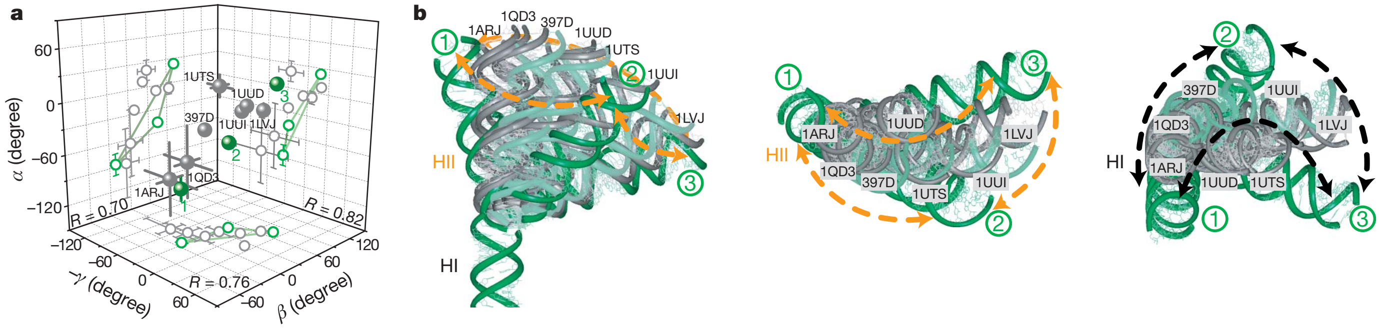


Figure 4 | Correlated TAR dynamics steer ligand-induced transitions. **a**, The three TAR dynamical conformers (green) and the TAR conformation (grey) bound to peptide derivatives of Tat (1ARJ)[16,17], divalent ions (397D)[18] and five different small molecules (1QD3, 1LVJ, 1UUD, 1UUI, 1UTS)[19–22]. Shown on each 2D plane is the correlation coefficient (R) between angles for the ligand-bound conformations. For bound structures, the s.d. is over the family of reported NMR models. Only one model was reported in structures 397D, 1UUI and 1UUD; thus no s.d. is reported. **b**, Comparison of the three TAR dynamical conformers (green) and ligand-bound TAR conformations (grey). Sub-conformers along the linear pathway linking conformers 1→2, 2→3 and 3→1 are shown in light green, and the direction of the trajectory is shown with arrows. Horizontal view following superposition of HI (left panel), vertical view down and up the helix axis of HI (middle panel) and HII (right panel) following superposition of HI and HII, respectively.

and pH approximately 6.4) at 298 K on an Avance Bruker 600 MHz NMR spectrometer equipped with a triple-resonance 5 mm cryogenic probe, and for the EI-TAR+ARG complex on a Varian Inova 800 MHz NMR spectrometer equipped with a 5 mm triple-resonance probe. The base ^{1}H–^{13}C splittings were measured from the difference between the upfield and downfield components of the ^{1}H–^{13}C doublet along the ^{1}H dimension using the narrow transverse relaxation-optimized spectroscopy (TROSY) component in the ^{13}C dimension as implemented in 2D ^{1}H–^{13}C S^{3}CT-heteronuclear single-quantum correlation (HSQC) experiments[30]. The ^{1}H–^{15}N splittings were measured in duplicate by using standard HSQC experiments without decoupling in the indirect or direct dimension. RDCs measured in E-AU-TAR and E-GC-TAR were normalized as detailed in Methods to take into account small differences in the degree of alignment arising from use of a slightly different Pf1 phage concentration[23]. The normalized RDCs were combined in the order-tensor analysis. Errors in the helix order tensors due to RDC uncertainty and A-form 'structural noise' were computed using the program AFORM-RDC[26]. Idealized A-form helices were constructed by using Insight II (Molecular Simulations, Inc.), noting that the propeller twist angles had to be corrected from +15° to the standard A-form value of −15° (ref. 26). The ensemble search was performed by using software written in-house as detailed in Methods. The Euler angles (α, β, γ) used in this work follow the y-convention and transform the HII frame from an alignment that is perfectly coaxial with HI, with the helix axis oriented along the molecular z direction to the final orientation in the conformer. Positive angles correspond to anti-clockwise rotations, and positive α and γ angles correspond to under- and over-twisting of HII and HI, respectively.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 19 July; accepted 18 October 2007.

1. Micura, R. & Hobartner, C. On secondary structure rearrangements and equilibria of small RNAs. *ChemBioChem* **4**, 984–990 (2003).
2. Schroeder, R., Barta, A. & Semrad, K. Strategies for RNA folding and assembly. *Nature Rev. Mol. Cell Biol.* **5**, 908–919 (2004).
3. Schwalbe, H. *et al.* Structures of RNA switches: insight into molecular recognition and tertiary structure. *Angew. Chem. Int. Edn Engl.* **46**, 1212–1219 (2007).
4. Leulliot, N. & Varani, G. Current topics in RNA-protein recognition: control of specificity and biological function through induced fit and conformational capture. *Biochemistry* **40**, 7947–7956 (2001).
5. Al-Hashimi, H. M. Dynamics-based amplification of RNA function and its characterization by using NMR spectroscopy. *ChemBioChem* **6**, 1506–1519 (2005).
6. Tjandra, N. & Bax, A. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **278**, 1111–1114 (1997).
7. Tolman, J. R., Flanagan, J. M., Kennedy, M. A. & Prestegard, J. H. NMR evidence for slow collective motions in cyanometmyoglobin. *Nature Struct. Biol.* **4**, 292–297 (1997).
8. Furtig, B. *et al.* Time-resolved NMR studies of RNA folding. *Biopolymers* **86**, 360–383 (2007).
9. Blackledge, M. Recent progress in the study of biomolecular structure and dynamics in solution from residual dipolar couplings. *Prog. Nucl. Magn. Reson. Spectrosc.* **46**, 23–61 (2005).
10. Mittermaier, A. & Kay, L. E. New tools provide new insights in NMR studies of protein dynamics. *Science* **312**, 224–228 (2006).
11. Palmer, A. G. III & Massi, F. Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. *Chem. Rev.* **106**, 1700–1719 (2006).
12. Bruschweiler, R. New approaches to the dynamic interpretation and prediction of NMR relaxation data from proteins. *Curr. Opin. Struct. Biol.* **13**, 175–183 (2003).
13. Zhang, Q., Sun, X., Watt, E. D. & Al-Hashimi, H. M. Resolving the motional modes that code for RNA adaptation. *Science* **311**, 653–656 (2006).
14. Zhang, Q. & Al-Hashimi, H. M. Extending the NMR spatial resolution limit by motional couplings. *Nature Methods* (submitted).
15. Aboul-ela, F., Karn, J. & Varani, G. Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. *Nucleic Acids Res.* **24**, 3974–3981 (1996).
16. Puglisi, J. D. *et al.* Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science* **257**, 76–80 (1992).
17. Aboul-ela, F., Karn, J. & Varani, G. The structure of the human-immunodeficiency-virus type-1 Tar RNA reveals principles of RNA recognition by Tat protein. *J. Mol. Biol.* **253**, 313–332 (1995).
18. Ippolito, J. A. & Steitz, T. A. A 1.3-angstrom resolution crystal structure of the HIV-1 trans- activation response region RNA stem reveals a metal ion-dependent bulge conformation. *Proc. Natl Acad. Sci. USA* **95**, 9819–9824 (1998).
19. Faber, C., Sticht, H., Schweimer, K. & Rosch, P. Structural rearrangements of HIV-1 Tat-responsive RNA upon binding of neomycin B. *J. Biol. Chem.* **275**, 20660–20666 (2000).
20. Du, Z., Lind, K. E. & James, T. L. Structure of TAR RNA complexed with a Tat–TAR interaction nanomolar inhibitor that was identified by computational screening. *Chem. Biol.* **9**, 707–712 (2002).
21. Davis, B. *et al.* Rational design of inhibitors of HIV-1 TAR RNA through the stabilisation of electrostatic 'hot spots'. *J. Mol. Biol.* **336**, 343–356 (2004).
22. Murchie, A. I. *et al.* Structure-based drug design targeting an inactive RNA conformation: exploiting the flexibility of HIV-1 TAR RNA. *J. Mol. Biol.* **336**, 625–638 (2004).
23. Hansen, M. R., Mueller, L. & Pardi, A. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature Struct. Biol.* **5**, 1065–1074 (1998).
24. Meiler, J. *et al.* Model-free approach to the dynamic interpretation of residual dipolar couplings in globular proteins. *J. Am. Chem. Soc.* **123**, 6098–6107 (2001).
25. Tolman, J. R., Al-Hashimi, H. M., Kay, L. E. & Prestegard, J. H. Structural and dynamic analysis of residual dipolar coupling data for proteins. *J. Am. Chem. Soc.* **123**, 1416–1424 (2001).
26. Musselman, C. *et al.* Impact of static and dynamic A-form heterogeneity on the determination of RNA global structural dynamics using NMR residual dipolar couplings. *J. Biomol. NMR* **36**, 235–249 (2006).
27. Saupe, A. Recent results in the field of liquid crystals. *Angew. Chem. Int. Edn Engl.* **7**, 97–112 (1968).
28. Clore, G. M. & Schwieters, C. D. Amplitudes of protein backbone dynamics and correlated motions in a small alpha/beta protein: correspondence of dipolar coupling and heteronuclear relaxation measurements. *Biochemistry* **43**, 10678–10691 (2004).
29. Al-Hashimi, H. M. *et al.* Concerted motions in HIV-1 TAR RNA may allow access to bound state conformations: RNA dynamics from NMR residual dipolar couplings. *J. Mol. Biol.* **315**, 95–102 (2002).

30. Meissner, A. & Sorensen, O. W. The role of coherence transfer efficiency in design of TROSY- type multidimensional NMR experiments. *J. Magn. Reson.* **139,** 439–442 (1999).
31. Cornilescu, G., Marquardt, J. L., Ottiger, M. & Bax, A. Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J. Am. Chem. Soc.* **120,** 6836–6837 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Kurochkin for NMR expertise. We acknowledge the Michigan Economic Development Cooperation and the Michigan Technology Tri-Corridor for the support of the purchase of a 600 MHz spectrometer, and the W. F. Keck Foundation, National Science Foundation (NSF) and National Institutes of Health (NIH) for funds for the purchase of an 800 MHz spectrometer. Supported by an NIH and NSF grant.

Author Contributions H.M.A. conceived the technique, Q.Z. and H.M.A. analysed the data and wrote the paper. Q.Z., C.K.F. and H.M.A. analysed the HIV-2 TAR data. Q.Z., A.C.S. and C.K.F. prepared the samples and performed the NMR experiments.

Author Information Supplementary Movies are hosted at the publicly accessible Database of Macromolecular Movements (http://www.molmovdb.org/). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to H.M.A. (hashimi@umich.edu).

METHODS

Normalization of RDCs. The RDCs (Supplementary Table 1) were measured independently for HIV-1 EI-AU-TAR, EI-GC-TAR, HIV-2 EI-AU-TAR and EI-GC-TAR by using approximately 8 mg ml^{-1} Pf1 phage and for HIV-1 EII-AU-TAR, EII-GC-TAR, EI-AU-TAR+ARG and EI-GC-TAR+ARG using approximately 6 mg ml^{-1} Pf1 phage ordering medium[23]. The E-AU-TAR and E-GC-TAR RDCs were normalized before combination in the order-tensor analysis by repeatedly fitting the RDCs to each helix after uniform scaling of the E-GC-TAR RDCs by a normalization factor L. For EI-TAR, EII-TAR, EI-TAR+ARG and HIV-2 EI-TAR, the two helices yielded a similar best-fit L value of 0.84 (0.82 for HI and 0.85 for HII), 0.97 (0.97 for both HI and HII), 0.75 (0.71 for HI and 0.79 for HII) and 1.01 (1.01 for both HI and HII), respectively. Insignificant variations were observed in the order-tensor analysis and the ensemble search when varying L by $\pm$0.05.

Computing inter-helix Euler angles. Euler angles for the bound TAR structures (Fig. 4a) were computed as follows. Idealized A-form helices were constructed by using Insight II (Molecular Simulations, Inc.) (propeller twist angles had to be corrected from $+15^\circ$ to -15°) and superimposed onto the two helices of the bound TAR structures (PDB ID 1ARJ, 397D, 1LVJ, 1QD3, 1UUD, 1UUI and 1UTS), excluding terminal residues 17 and 45 when superimposing helices to HI. The inter-helix Euler angles (α, β, γ) were then computed by using EulerRNA[32] and a molecular frame in which the HI helix was oriented along the z direction with the 5′ and 3′ ends along the negative and positive z direction, respectively.

Computing axis of asymmetry for ligand-bound conformations. The S_{yy} principal direction (Fig. 2c) was computed as follows. Idealized A-form helices were superimposed onto the helices of the bound TAR structures as described above. Synthetic RDCs were computed for each helix by using the experimental order tensor determined for the elongated helices (Supplementary Table 2). The synthetic short helix RDCs were then averaged over the seven bound structures assuming equal populations and used to determine order tensors for each helix by using the same procedure used to determine the experimental order tensors shown in Fig. 2c.

Ensemble search. The ensemble search was performed by using in-house software. Trial inter-helical orientations defined by the inter-helix Euler angles α, β, γ were generated, and RDCs computed for each of the two short helices using the order tensors determined for their corresponding elongated helices (Supplementary Table 2). The predicted RDCs were then compared with measured counterparts, either directly ($N=1$) or after 'dynamical' averaging over two ($N=2$) or more ($N\geq 3$) trial conformations. Specifically, Q factors[31] were computed for each helix (Q_{HI} and Q_{HII}) and for the helices combined $Q^{\mathrm{tot}}=((Q^2_{\mathrm{HI}}+Q^2_{\mathrm{HII}})/2)^{1/2}$ using the degree of order for elongated helices in the normalization:
$$Q_{\mathrm{HI(HII)}}=\left(\sum_i\left(\mathrm{RDC}_i^{\mathrm{meas}}-\mathrm{RDC}_i^{\mathrm{pred}}\right)^2/\sum_i\left(D^2_{\max,i}S^2_{zz\mathrm{EII(EI)}}(3+\eta^2_{\mathrm{EII(EI)}})/15\right)\right)^{1/2},$$
$$D_{\max}=-\frac{\mu_0\gamma_m\gamma_n h}{8\pi^3 r^3_{m,n}} \text{ and } \eta=\frac{S_{xx}-S_{yy}}{S_{zz}}.$$

The distance between C39 O3′ and U40 P was maintained at 1.60 Å. Loose steric constraints were imposed by excluding conformations that lead to: (i) van der Waals collisions between the two A-form helices, using uniform and conservative van der Waals radii of 2.5 Å excluding protons as well as all the terminal A22-U40 base pair (the latter was used to allow deformations in this more flexible region of the A-form structure to take place); and (ii) a distance between A22 O3′ and G26 P that is greater than the theoretically allowed length of the trinucleotide bulge (21 Å)[33]. The ensemble search was conducted in two stages. In the first stage, an exhaustive grid search over the 3- ($N=1$), 6- ($N=2$) or 9- ($N=3$) dimensional α, β, γ parameter space was performed using step-size angles of $\alpha=20^\circ$, $\beta=10^\circ$ and $\gamma=20^\circ$. Solutions with $Q_{\mathrm{HI}}\leq Q^{\mathrm{I}}$ and $Q_{\mathrm{HII}}\leq Q^{\mathrm{II}}$ were subjected to a second round of refinement. Q^{I} and Q^{II} were determined by Monte-Carlo-type simulations. Briefly, the average Q value was computed after the imposition of (i) inter-helix 'structural noise' in each conformer consisting of rotations with step-size angles $\alpha=20^\circ$, $\beta=10^\circ$, $\gamma=20^\circ$, and (ii) RDC uncertainty (about 3.4 Hz and about 1.9 Hz for EI-TAR and EII-TAR, respectively; note the smaller RDC error in EII-TAR is attributed to its approximate 41% smaller degree of alignment compared with EI-TAR (Supplementary Table 2)). For $N=1$, all solutions were subjected to a second round of refinement. For $N>1$, 10% of the accepted ensembles yielding the lowest Q^{tot} values and another randomly chosen 10% were subjected to further rounds of refinement. In the second stage, each ensemble was further refined by performing a narrower search over its three conformers. This was accomplished by perturbing the original inter-helix Euler α, β, γ angles defining each conformer by $\alpha=\pm 20^\circ$, $\beta=\pm 10^\circ$, $\gamma=\pm 20^\circ$ in increments of 10°, 5° and 10°, respectively. The best-fit solution yielding the lowest Q^{tot} value was then subjected to two additional rounds of refinement involving increasingly smaller perturbations: $\alpha=\pm 10^\circ$, $\beta=\pm 5^\circ$, $\gamma=\pm 10^\circ$, in increments of 5°, 2.5° and 5°, respectively; and finally $\alpha=\pm 5^\circ$, $\beta=\pm 2.5^\circ$, $\gamma=\pm 5^\circ$, in increments of 2.5°, 1.25° and 2.5°, respectively. Final ensemble solutions were accepted if both Q_{HI} and $Q_{\mathrm{HII}}\leq Q^{\mathrm{final}}$, where Q^{final} takes into account A-form structural noise and RDC uncertainty ($Q=7.5\%$), uncertainty in the order tensors determined for the elongated helices ($Q=6.2\%$) and finite step size in the final round of refinement ($Q=2.9\%$). For $N=3$, the value of Q^{final} was 10.2%. The performance of this protocol was validated by extensive simulations. The final solutions were clustered into ensembles A (40%), B (48%) and C (12%). The minor solution C could easily be excluded because in one conformer the two helices could not be linked by the trinucleotide bulge without causing steric collisions.

Cross-validation of the dynamical ensembles. Order tensors were predicted for each of the three TAR conformers in ensembles A and B by using Prediction of ALignmEnt from Structure (PALES)[34]. The order tensors for each helix were then averaged over the three conformers and resulting values compared with experimental counterparts[29]. Although solution A yields very good agreement between the predicted and measured helix order-tensor frames (Supplementary Fig. 5a) and RDCs (Supplementary Fig. 5a), the agreement is significantly less favourable for solution B (Supplementary Fig. 5b). The predicted S_{zz} deviates by 24° and 13° for HII and HI, respectively, and less favourable RDC agreement ($R=0.89$) is observed (Supplementary Fig. 5b). Similarly, poor agreement was obtained for a motional model involving inter-conversion among all six conformers (data not shown).

Measurement and analysis of ^{15}N spin relaxation data in HIV-2 EI-TAR. Imino ^{15}N longitudinal (R_1), transverse ($R_{2(\mathrm{CPMG})}$) relaxation rates and $\{^1\mathrm{H}\}$-^{15}N nuclear Overhauser effects (NOEs) were measured as described previously[13]. The relaxation delays for R_1 and $R_{2(\mathrm{CPMG})}$ experiments were 0.06, 0.12 (×2), 0.24, 0.48, 0.64, 0.80 (×2) and 1.2 s, and 0.0062, 0.0124 (×2), 0.0248, 0.0372, 0.0496, 0.0620 (×2) and 0.0744 s, respectively (duplicate measurements are denoted by '×2') (Supplementary Fig. 4a). Data were analysed by using the extended model-free formalism[35] (Supplementary Table 3) as implemented in Modellfree (Version 4.16 for Linux) from Palmer and co-workers[36], as described previously[13]. The RDC-derived HIV-2 EI-TAR inter-helix orientation was used as input structure for the model-free analysis.

32. Bailor, M. H. *et al.* Characterizing the relative orientation and dynamics of RNA A-form helices using NMR residual dipolar couplings. *Nature Protoc.* **2,** 1536–1546 (2007).
33. Saenger, W. *Principles of Nucleic Acid Structure* (Springer-Verlag, New York, New York, 1984).
34. Zweckstetter, M. & Bax, A. Prediction of sterically induced alignment in a dilute liquid crystalline phase; aid to protein structure determination by NMR. *J. Am. Chem. Soc.* **122,** 3791–3792 (2000).
35. Clore, G. M. *et al.* Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc.* **112,** 4989–4991 (1990).
36. Mandel, A. M., Akke, M. & Palmer, A. G. Backbone dynamics of *Escherichia coli* ribonuclease Hi: correlations with structure and function in an active enzyme. *J. Mol. Biol.* **246,** 144–163 (1995).

LETTERS

Segrosome structure revealed by a complex of ParR with centromere DNA

Maria A. Schumacher[1], Tiffany C. Glover[1], Anthony J. Brzoska[2], Slade O. Jensen[2], Thomas D. Dunham[1], Ronald A. Skurray[2] & Neville Firth[2]

The stable inheritance of genetic material depends on accurate DNA partition. Plasmids serve as tractable model systems to study DNA segregation because they require only a DNA centromere, a centromere-binding protein and a force-generating ATPase. The centromeres of partition (*par*) systems typically consist of a tandem arrangement of direct repeats[1–7]. The best-characterized *par* system contains a centromere-binding protein called ParR and an ATPase called ParM. In the first step of segregation, multiple ParR proteins interact with the centromere repeats to form a large nucleoprotein complex of unknown structure called the segrosome, which binds ParM filaments[4,8–10]. pSK41 ParR binds a centromere consisting of multiple 20-base-pair (bp) tandem repeats to mediate both transcription autoregulation and segregation. Here we report the structure of the pSK41 segrosome revealed in the crystal structure of a ParR–DNA complex. In the crystals, the 20-mer tandem repeats stack pseudo-continuously to generate the full-length centromere with the ribbon–helix–helix (RHH) fold of ParR binding successive DNA repeats as dimer-of-dimers. Remarkably, the dimer-of-dimers assemble in a continuous protein super-helical array, wrapping the DNA about its positive convex surface to form a large segrosome with an open, solenoid-shaped structure, suggesting a mechanism for ParM capture and subsequent plasmid segregation.

Plasmid partition is performed by one of three characterized types of *par* system. Type I systems use ParA ATPase proteins with deviant Walker-type folds and centromere-binding proteins called ParB; type II systems use actin-like ParM ATPases and centromere-binding proteins called ParR; and a recently described type III system uses a tubulin-like protein, TubZ[1–7, 11]. The clinically important multidrug-resistance plasmid pSK41 contains a putative type II *par* system within a 1.6-kilobase (kb) region (Supplementary Information)[9]. Plasmid segregational stability assays, which showed that a plasmid carrying the pSK41 *par* region was maintained by 83% of cells compared with 28% with the vector alone after 120 generations, demonstrated that the pSK41 *par* region represents a bone fide *par* system. Site-directed mutants that either abolished function of ParM (D190A) or removed the ParR protein (K4 to stop codon) eliminated the plasmid stability phenotype, confirming that both ParR and ParM are required for partition (Supplementary Information; Supplementary Fig. 1). Furthermore, studies of reporter gene fusions revealed that, like the homologous ParR from the prototypical R1 type II *par* system[5–10], pSK41 ParR autoregulates the pSK41 *parMR* operon (Supplementary Table 1; Supplementary Fig. 2a, b). The autoregulation function of ParR also explains the hyper-instability phenotype associated with the *parR* nonsense mutant in plasmid segregation assays (Supplementary Fig. 1), as loss of the ParR auto-repression activity leads to uncontrolled expression of toxic Par proteins.

Centromeres are typically located upstream or downstream of the *par* genes[4,7]. Consistent with this, DNase I protection studies revealed that pSK41 ParR protects a region directly upstream of the *par* operon (Fig. 1a, b). This centromere consists of a series of tandem DNA repeats that can be divided into eight 10-bp or four 20-bp repeats (Fig. 1a). The DNase I protection studies show a striking degree of periodicity, indicative of highly symmetric interactions of multiple ParR molecules with the centromere. The arrangement of the pSK41 centromere is somewhat similar to that of R1, which consists of ten 11-bp direct repeats. However, unlike the pSK41 centromere, the R1 centromere repeats are not contiguous, but are

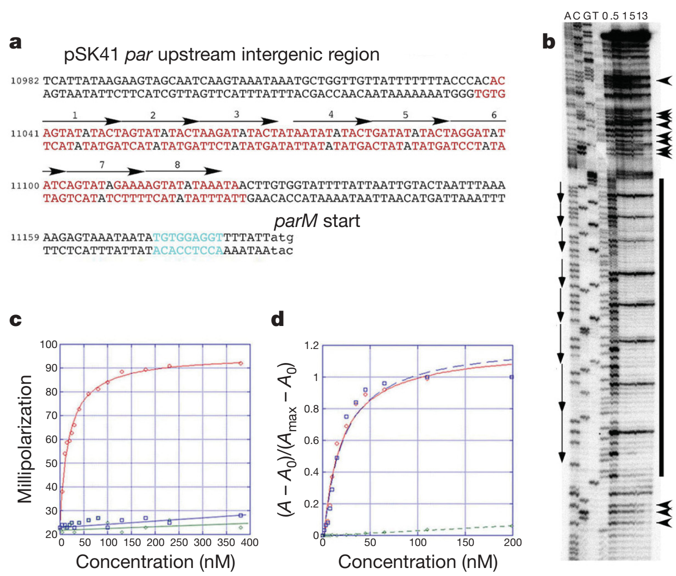


Figure 1 | The pSK41 centromere-like site. **a**, pSK41 centromere sequence. Sequences protected from DNase I are highlighted in red, positions of DNase-I-susceptible nucleotides Ade5 and Ade6 are black. **b**, DNase I footprint of ParR bound to *parC*. Vertical bar shows the protected region. Spreading is suggested by additional protection (arrowheads) at high ParR concentrations. Arrows denoting repeat positions are indicated on the sequencing ladder. **c**, Fluorescence polarization of ParR binding to centromere and non-centromere DNA. Fluorescence polarization units (mP, millipolarization) and ParR concentration are along the *y* and *x* axes, respectively. ParR interactions with: 10-bp centromere repeat (squares); 20-bp repeat (circles); 20-bp non-centromere site (diamonds). **d**, Fluorescence polarization of full-length ParR, ParRN and ParRC binding to 20-bp repeat. Normalized fluorescence polarization ($(A - A_0)/(A_{max} - A_0)$) and protein concentration are along the *y* and *x* axes, respectively. Interactions of 20-mer with: full-length ParR (squares); ParRN (circles); ParRC (diamonds).

[1]Department of Biochemistry and Molecular Biology, University of Texas, M.D. Anderson Cancer Center, Unit 1000, Houston TX 77030, USA. [2]School of Biological Sciences, University of Sydney, Sydney, New South Wales 2006, Australia.

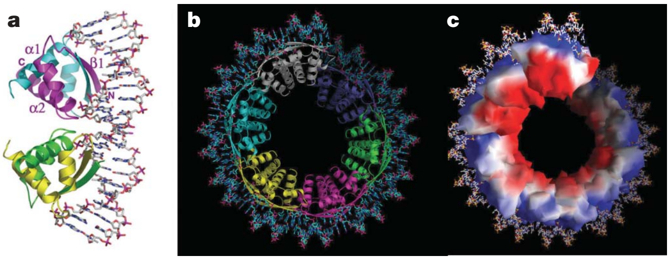


Figure 2 | Structure of pSK41 ParR–DNA segrosome. **a**, Overall structure of the ParR–DNA complex. ParRN subunits are coloured magenta, cyan, green and yellow. DNA carbon, nitrogen, oxygen and phosphorus atoms are coloured white, blue, red and magenta, respectively. This figure and Figs 2b, 3a, 3c and 4b were made with PyMOL[28]. **b**, Ribbon diagram of the ParRN–DNA segrosome complex. The view is looking down the super-helix axis. ParRN dimer-of-dimers are white, cyan, yellow, magenta, green and blue. **c**, Electrostatic surface representation of the segrosome complex, shown in the same orientation as Fig. 2b. Blue and red represent electropositive and electronegative surfaces, respectively. This figure was made using GRASP[29].

divided into two sets of five repeats located on either side of the promoter[7]. Thus, the pSK41 centromere represents one of the simplest *par* centromeres that has been described. To elucidate the minimal direct repeat required for binding by pSK41 ParR, a fluorescence polarization assay was used[12]. ParR bound neither the 10-bp repeat nor a non-centromere control site; however, it bound the 20-bp repeat with high affinity ($K_d = 17.5 \pm 4.3$ nM), indicating that the 20-bp site is the minimal centromere repeat (Fig. 1c). pSK41 ParR can be digested into an amino (N) domain, residues 1–53 (ParRN), and a carboxy (C) domain, residues 54–109 (ParRC). Fluorescence polarization revealed that the ParRN domain contains all the determinants required for centromere binding as ParRN bound the 20-mer with a K_d essentially identical to that of full-length ParR, whereas ParRC showed no binding (Fig. 1d). Studies have shown that ParR/ParB binding to centromere sites leads to the creation of higher-order nucleoprotein complexes, called segrosomes, the formation of which involves DNA wrapping[6]. To gain insight into how type II ParR recognizes its centromere and forms a segrosome, we performed crystallographic studies. Data-quality crystals were obtained with ParRN bound to the 20-mer repeat. The crystallographic asymmetric unit contains one ParRN dimer and one 20-mer half-site, and the structure has a value of R_{work}/R_{free} of 25.5%/29.6% to 3.0 Å resolution (Supplementary Table 2)[13,14]. Crystallographic symmetry generates the ParRN dimer-of-dimer/20-mer duplex as well as the entire pSK41 segrosome (Fig. 2a–c).

ParRN contains an RHH DNA-binding fold, which is a common nucleic-acid-binding motif found in the Arc/MetJ family of DNA-binding proteins[15–21]. Notably, this fold is also present in the type I centromere-binding proteins ParG and ω[17,21]. Structure homology searches show that the ParR RHH has highest similarity to that of the *Escherichia coli* proline utilization A protein (PutA)[20]. The ParRN RHH topology is: β1, residues 5–12; α1, residues 16–25; α2, residues 33–47 (Fig. 2a). Like other RHH proteins, the ParRN β-strands combine in an antiparallel fashion and the α-helices interdigitate to form a tight dimer. The dimer buries 3678 Å^2 of protein surface area from solvent, which corresponds to approximately 65% of its total surface area. However, the most notable feature of the ParRN–20-mer structure is the creation, in the crystal, of the pSK41 segrosome whereby the 20-mer DNA elements are packed pseudo-continuously to create a full-length centromere. The protein dimers that bind the DNA form intimate protein–protein interactions to generate a continuous protein super-structure that has distinctly positive and negative surfaces (Fig. 2b, c). The positive surface of the ParR super-structure wraps the DNA about its convex surface to create a unique super-helical structure, called the segrosome. Characteristic of the structure is its super-helical parameters; it has a pitch of about 240 Å with six ParRN dimer-of-dimers in one turn of the super-helix and a diameter of about 178 Å. The large diameter of the segrosome indicates that it has a much more extended conformation compared with other known protein–DNA super-helical structures such as the core nucleosome[22].

The DNase I footprint data of ParR binding to the centromere reveal nucleosome-like periodicity within the ParR-binding region consistent with the crystal structure (Fig. 1b). Specifically, the footprint shows almost continuous protection of the centromere when bound by ParR, which would be expected from a contiguous protein super-helix such as found in the crystal structure. In addition, eight Ade nucleotides (on each strand), corresponding to Ade5 on one strand and Ade6 on the opposite strand (Figs 1b and 3a, b), which are symmetrically distributed along the protected region, are susceptible to DNase I attack. The basis for the susceptibility of these nucleotides is revealed by the crystal structure, which shows that they reside in an exposed minor groove position between ParR molecules that bind each repeat (Fig. 3a, b). The structure also explains why there are no other sites susceptible to DNase I: DNase I acts in the minor groove, and in the structure only the minor groove side containing nucleotides 5 and 6 is accessible as the opposite minor groove face is embraced by ParR molecules and buried within the centre of the segrosomal super-structure (Supplementary Fig. 3). Moreover, because DNase I requires approximately four to five nucleotides upstream and downstream of its cleavage site, only the centrally located nucleotides 5 and 6 are susceptible in the complex, as DNase I cannot effectively dock onto upstream or downstream

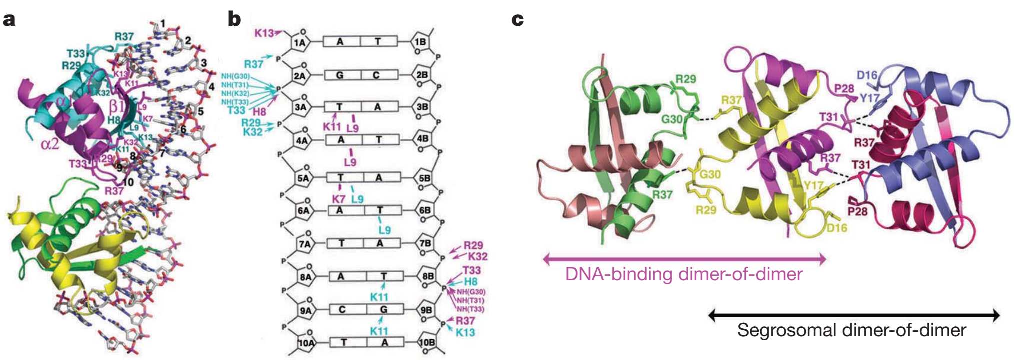


Figure 3 | ParR–DNA and ParR–ParR interactions in the segrosome. **a**, Ribbon diagram showing interactions of one ParR dimer with its DNA-site. Contacts made to each 10-bp half-site are identical in the crystal structure, thus only contacts to one half-site are shown. **b**, Schematic diagram of ParRN–DNA contacts. The riboses of each nucleotide are numbered and shown as pentagons. Contacts are colour-coded according to the interacting subunit. Hydrogen bonds and hydrophobic interactions are shown as arrows and lines, respectively, from residue to base. **c**, Interface interactions. The DNA-binding dimer-of-dimer and the segrosomal super-helix dimer-of-dimer interfaces are shown. Residues participating are shown as sticks; hydrogen bonds are indicated by dashed lines.

regions without encountering bound ParR molecules. Thus, the multiple structural features revealed in the DNase I footprint are all supported by the crystal structure. However, to further test the crystal structure model, we performed electron microscopy on ParR bound to the full-length centromere site. Electron microscopy images of ParR–centromere complexes revealed circular structures with dimensions essentially identical to the super-helical dimensions found in the crystal structure (Supplementary Fig. 4). Thus, DNase I and electron microscopy studies support the segrosome structure observed in the crystal; together, these data provide a structural basis for previously described ParR-mediated DNA wrapping[6].

ParRN binds the 20-bp site as a dimer-of-dimers whereby each dimer contacts 10-bp of the repeat. Although the DNA in the ParRN–centromere complex has mainly B-form characteristics (average rise and twist of 3.38 Å and 33.0° compared with 3.38 Å and 34.3° for B-form DNA) significant distortions are seen including a global bend of 46° and an overall enlarged major groove (from 11.0 Å to about 14.0 Å) within each $T_3A_4T_5A_6$ stretch[23]. Phosphate contacts cluster on the 5′ ends of each 10-bp repeat and contribute to DNA distortions (Fig. 3a, b). These phosphate contacts represent the points of closest physical association of ParR with DNA, which again leaves the central nucleotides, Ade5 and Ade6, prone to DNase I attack (Figs 1b and 3a, b). The widening of the major groove allows insertion of the ParR antiparallel β-sheet and residues Lys7, Leu9 and Lys11, which make base contacts to nucleotides 3–6, 8 and 9. The clustering of the protein–base contacts to base pairs 3–6 is consistent with the fact that these are the most conserved nucleotides of the centromere repeats (Figs 1a and 3a, b). Indeed, only contacts to base pairs 3 and 6 are clearly base specific. Indirect readout may also play some role in centromere recognition as TATA sequences, which are most distorted in the structure, have the highest conformational flexibility of DNA sequences[24]. Once nucleated onto their centromere sites, ParR/ParB proteins coat the surrounding DNA, and our DNase I protection data show that pSK41 ParR also spreads to DNA sites surrounding the centromere at high protein concentrations (Fig. 1b)[1–5,25].

The segrosome structure suggests protein–DNA and protein–protein interactions arise cooperatively to generate the super-helical structure. In agreement, Hill plot analysis of electrophoretic mobility shift assay data confirmed that ParR binding to the centromere is highly cooperative ($n_H = 4.1$) (Supplementary Fig. 5). Interactions between ParR molecules that generate the segrosome are likely nucleated when the first ParR dimer-of-dimers binds and distorts its DNA site. The interface between the DNA-binding dimer-of-dimers buries approximately 360 Å^2 of protein surface from solvent and is stabilized by dyad-related hydrogen bonds between Arg 37 and the carbonyl oxygen of Arg 29 located on a subunit in the adjacent dimer (Fig. 3c). The interface that generates the segrosome is slightly more extensive than the DNA-binding interface, although it is formed from the same regions. This interface buries about 400 Å^2 of protein surface and is mediated by hydrogen bonds between Thr31 on one dimer and residues Arg 37′ and Tyr 17 (where ′ indicates the other subunit of the dimer) from the subunits of the opposing dimer (Fig. 3c). Clearly, additional stabilization of these dimer-of-dimer interfaces may be imparted from C-domain residues not present in ParRN. However, the location of the exposed C domain in the segrosome suggests that it may function in ParM recruitment. This was confirmed by fluorescence polarization, which showed that ParM bound only to complexes containing full-length ParR and not ParRN (Fig. 4a; Supplementary Fig. 6).

A critical role of the segrosome is to recruit ParM filaments and prevent them from undergoing catastrophic collapse, ultimately allowing filament growth and plasmid separation[9]. Our segrosome structure suggests a mechanism for ParM capture. Specifically, the structure shows that even the minimal pSK41 centromere, containing four ParR dimer-of-dimer 20-bp repeat units and thus two-thirds of a super-helical turn, would contain a central region of exposed and concentrated ParR C termini poised for ParM capture. Moreover, the segrosome diameter is about 18 nm, which is large enough to capture ParM filaments (about 6 nm in diameter) inside or, if the C domains fill the pore, on the surface of its central core (Fig. 4b). ParR spreading could help further encase ParM molecules[1–5,25]. Could this proposed segrosome capture mechanism apply to other *par* systems? ParR/ParB proteins show little sequence identity. However, recent structures have started to reveal similarities in the centromere-binding elements of these proteins. For example, several of the larger type I ParB structures contain helix–turn–helix (HTH) motifs, whereas smaller type I ParB proteins contain RHH folds[17,21,26,27]. The structure reported here adds an unexpected twist to these findings as it reveals that type II ParR also contains an RHH. In fact, the RHH fold is an excellent segrosome-forming motif because binding to tandem centromere repeats would promote its cooperative polymerization and formation of super-structures. Whether this type of polymerization is of general importance in segrosome formation remains to be seen. However, recently identified similarities between *par* systems, such as formation of filaments by type I ParA proteins, suggests that *par* systems likely use similar segregation mechanisms, which may include the formation of segrosomal super-structures.

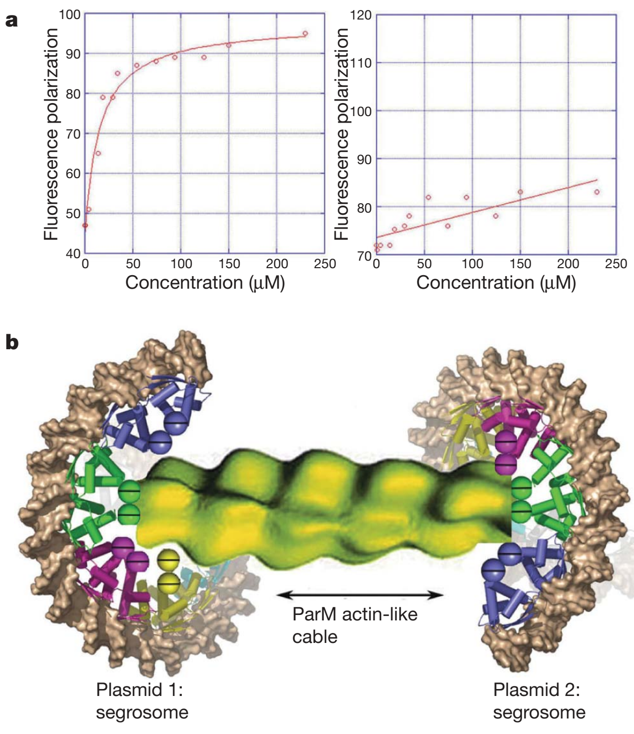


Figure 4 | pSK41 segregation model. a, Fluorescence polarization identifying ParM-interacting domain within ParR. Isotherms are for ParM-(AMP-PCP)-Mg^{2+} titration into full-length ParR–DNA (left) and ParRN–DNA complexes (right). ParM-(AMP-PCP)-Mg^{2+} binds only the full-length ParR–centromere complex (the apparent K_d is 12 ± 3 μM). **b**, pSK41 segregation model. ParM filament capture between two segrosomes on different plasmids is shown with molecules drawn roughly to scale. Segrosomal DNA is coloured brown, the ParRN molecules magenta, green, blue and yellow, and ParR molecules participating in spreading are cyan. ParR C domains are represented as half circles. The ParM filament is represented as a yellow surface rendering[30].

METHODS SUMMARY

Protein purification. pSK41 *parR* and *parM* genes were subcloned into pTTQ18 with C-terminal encoding RGSHHHHHH tags added for purification. The proteins were purified using Ni-NTA chromatography. To identify ParR domains, the full-length protein was digested with 0.1% trypsin. After 2 h, two major regions were identified: a stable N-terminal domain (residues 1–53) and a C-terminal region (residues 54–109), which was less stable than ParRN to further proteolysis. These domains were subcloned in pET-15b, expressed and purified.

Fluorescence polarization studies. Fluorescence polarization measurements were collected with a PanVera Beacon 2000 Fluorescence Polarization System. Samples were excited at 490 nM and fluorescence emission was measured at 520 nM. Details of fluorescence polarization experiments used to elucidate the minimal centromere-binding site, the minimal ParR centromere-binding domain, and studies on ParM binding to pre-formed ParR–DNA and ParRN–DNA complexes, are given in Methods.

ParR–DNA crystallization, data collection and structure determination. For crystallization, centromere sites of 20 bp or longer were used. None of the full-length ParR–DNA crystals diffract beyond 35.0 Å resolution. Therefore, co-crystallization trials were done with ParRN. Data-quality crystals were obtained of a ParRN–20-mer complex using a 2:1 ratio of ParRN dimer:20-mer duplex and a crystallization solution of 20% PEG 4000, 10% isopropanol and 0.1 M HEPES pH 7.5. Intensity data were collected at ALS beamline 8.3.1 at 100 K and processed with MOSFLM. The structure was solved by single isomorphous replacement (SIR); the final model included residues 4–47 of one ParRN subunit, 4–48 of the other, the 20-mer half-site, and was refined to R_{work}/R_{free} of 25.5%/29.6% to 3.0 Å resolution (Supplementary Table 2).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 July; accepted 18 September 2007.

1. Møller-Jensen, J., Jensen, R. B. & Gerdes, K. Plasmid and chromosome segregation in prokaryotes. *Trends Microbiol.* **8**, 313–320 (2000).
2. Schumacher, M. A. Structural biology of plasmid segregation proteins. *Curr. Opin. Struct. Biol.* **17**, 103–107 (2007).
3. Gerdes, K., Møller-Jensen, J., Ebersbach, G., Kruse, T. & Nordström, K. Bacterial mitotic machineries. *Cell* **116**, 359–366 (2004).
4. Hayes, F. & Barillà, D. The bacterial segrosome: a dynamic nucleoprotein machine for DNA trafficking and segregation. *Nat. Rev. Microbiol.* **4**, 133–143 (2006).
5. Jensen, R. B. & Gerdes, K. Partitioning of plasmid R1. The ParM protein exhibits ATPase activity and interacts with the centromere-like ParR-*parC* complex. *J. Mol. Biol.* **269**, 505–513 (1997).
6. Jensen, R. B., Lurz, R. & Gerdes, K. Mechanism of DNA segregation in prokaryotes: replicon pairing by *parC* of plasmid R1. *Proc. Natl Acad. Sci. USA* **95**, 8550–8555 (1998).
7. Dam, M. & Gerdes, K. Partitioning of plasmid R1. Ten direct repeats flanking the *parA* promoter constitute a centromere-like partitioning site *parC*, that expresses incompatibility. *J. Mol. Biol.* **236**, 1289–1298 (1994).
8. Garner, E. C., Campbell, C. S., Weibel, D. B. & Mullins, R. D. Reconstitution of DNA segregation driven by assembly of a prokaryotic actin homolog. *Science* **315**, 1270–1274 (2007).
9. Møller-Jensen, J., Jensen, R. B., Löwe, J. & Gerdes, K. Prokaryotic DNA segregation by an actin-like filament. *EMBO J.* **21**, 3119–3127 (2002).
10. Møller-Jensen, J. *et al.* Bacterial mitosis: ParM of plasmid R1 moves plasmid DNA by an actin-like insertional polymerization mechanism. *Mol. Cell* **12**, 1477–1487 (2003).
11. Larsen, R. A. *et al.* Treadmilling of a prokaryotic tubulin-like protein, TubZ, required for plasmid stability. *Genes Dev.* **21**, 1340–1352 (2007).
12. Lundblad, J. R., Laurance, M. & Goodman, R. H. Fluorescence polarization of protein–DNA and protein–protein interactions. *J. Mol. Endocrinol.* **10**, 607–612 (1996).
13. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
14. Brünger, A. T. *et al.* Crystallography and NMR System: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
15. Raumann, B. E., Rould, M. A., Pabo, C. O. & Sauer, R. T. DNA recognition by β-sheets in the Arc repressor–operator crystal structure. *Nature* **367**, 754–757 (1994a).
16. Raumann, B. E., Brown, B. M. & Sauer, R. T. Major groove DNA recognition by β-sheets: the ribbon–helix–helix family of gene regulatory proteins. *Curr. Opin. Struct. Biol.* **4**, 36–43 (1994b).
17. Weihofen, W. A., Cicek, A., Pratto, F., Alonso, J. C. & Saenger, W. Structures of ω repressors bound to direct and inverted DNA repeats explain modulation of transcription. *Nucleic Acids Res.* **34**, 1450–1458 (2006).
18. Somers, W. S. & Phillips, S. E. Crystal structure of the *met* repressor–operator complex at 2.8 Å resolution reveals DNA recognition by β-strands. *Nature* **359**, 387–393 (1992).
19. He, Y. Y. *et al.* Probing the *met* repressor–operator recognition in solution. *Nature* **359**, 431–433 (1992).
20. Larson, J. D. *et al.* Crystal structures of the DNA-binding domain of *Escherichia coli* proline utilization A flavoprotein and analysis of the role of Lys9 in DNA recognition. *Protein Sci.* **15**, 2630–2641 (2006).
21. Golovanov, A. P., Barillà, D., Golovanova, M., Hayes, F. & Lian, L. Y. ParG, a protein required for active partition of bacterial plasmids, has a dimeric ribbon–helix–helix structure. *Mol. Microbiol.* **50**, 1141–1153 (2003).
22. Luger, K., Mäder, A., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
23. Ravishanker, G., Swaminathan, S., Beveridge, D. L., Lavery, R. & Sklenar, H. Conformational and helicoidal analysis of 30 PS of molecular dynamics on the d(CGCGAATTCGCG) double helix: 'curves', dials and windows. *J. Biomol. Struct. Dyn.* **6**, 669–699 (1998).
24. Zhang, Y., Xi, Z., Hedge, R. S., Shakked, Z. & Crothers, D. M. Predicting indirect readout effects in protein–DNA interactions. *Proc. Natl Acad. Sci. USA* **101**, 8337–8341 (2004).
25. Rodionov, O., Lobocka, M. & Yarmolinsky, M. Silencing of genes flanking the P1 plasmid centromere. *Science* **283**, 546–549 (1999).
26. Schumacher, M. A. & Funnell, B. E. Structures of ParB bound to DNA reveal mechanism of partition complex formation. *Nature* **438**, 516–519 (2005).
27. Khare, D., Ziegelin, G., Lanka, E. & Heinemann, U. Sequence-specific DNA binding determined by contacts outside the helix–turn–helix motif of the ParB homolog KorB. *Nat. Struct. Mol. Biol.* **11**, 656–663 (2004).
28. Delano, W. L. *The PyMOL Molecular Graphics System* (DeLano Scientific, San Carlos, California, 2002).
29. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).
30. van den Ent, F., Møller-Jensen, J., Amos, L. A., Gerdes, K. & Löwe, J. F-actin-like filaments formed by plasmid segregation protein ParM. *EMBO J.* **21**, 6935–6943 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Advanced Light Source (ALS) and their support staff. The ALS is supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division of the US Department of Energy at the Lawrence Berkeley National Laboratory. We also thank Dr S. Kwong for supplying shuttle vectors and Professor K. Gerdes for pointing out the presence of the pSK41 *par* system. This work was supported by a Burroughs Wellcome Career Development Award, a U.T. M.D. Anderson Trust Fellowship and a National Institutes of Health grant (to M.A.S.), an Australian Research Council Grant (to N.F. and R.A.S.) and a National Health and Medical Research Council (Australia) Project Grant (to R.A.S. and N.F.).

Author Contributions M.A.S. performed the crystallographic studies, fluorescence polarization, oversaw cryo-electron microscopy studies and wrote the manuscript. T.C.G. generated ParRN and ParRC constructs. T.D.D. performed cryo-electron microscopy studies. A.J.B. generated plasmid constructs and undertook the regulatory and functional studies. S.O.J. performed the electrophoretic mobility shift assays (EMSA) and footprinting assays and contributed to the manuscript. R.A.S. and N.F. conceived and oversaw the functional studies and contributed to the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.A.S. (maschuma@mdanderson.org).

METHODS

Plasmid stability assays. *Staphylococcus aureus* RN4220 strains containing plasmids to be assayed were grown overnight in LB medium with the appropriate selection. Dilutions were prepared in saline, a viable count performed using non-selective LB agar plates, and the culture reinoculated (10^{-4} dilution) into 10 ml LB medium without selection. After overnight growth, the culture was diluted, counted and subcultured as before. This process was repeated until approximately 120 generations of growth were achieved. Fifty to 100 colonies from viable count plates were patched onto media with and without selection for the plasmid so that the proportion of the population retaining the resistance phenotype conferred by the plasmid could be quantified.

pSK9001 contains the 1.8 kb pSK41 *par* region cloned into the *E. coli–S. aureus* shuttle vector pSK5484; pSK9017 and pSK9018 are mutant derivatives of pSK9001. pSK9017 encodes a ParM protein with an aspartic acid to alanine substitution at position 190 (ParM-D190A). This residue is located in a conserved motif in the actin-like ATPases, including the ParM proteins, that is involved in Mg^{2+} binding[31]. An equivalent mutation in the R1 ParM protein abolishes function[5]. Sequence homology indicates that pSK41 ParM is structurally similar to R1 ParM. In fact, all proteins in the actin-like ATPase superfamily show a high degree of structure homology, and the residue corresponding to aspartic acid 190 in pSK41 ParM is conserved and critical for function[32]. pSK9018 contains an ochre termination codon in place of lysine 4 in ParR.

Chloramphenicol acetyltransferase assays. Chloramphenicol acetyltransferase (CAT) assays were performed in a microplate format as described previously[33]. CAT units are expressed as nanomoles of chloramphenicol acetylated per milligram of protein per minute at 37 °C, and are the average of at least three independent assays.

DNase I protection and EMSA studies. DNase I footprinting and EMSA were undertaken essentially as described previously[33]. ParR binding to *parC* was analysed using a 298-bp pSK41 DNA fragment (nt 10892–11190, GenBank entry AF051917), that was end-labelled using [γ-^{32}P]ATP. The amount of DNA probe bound by ParR was quantified by image analysis using QUANTITY ONE software (BioRad), and a Hill plot was constructed by graphing log(X/1 − X), where X is the fraction of bound fragment, against log(ParR concentration). The Hill coefficient (n_H) was given by the maximal slope; a value of greater than 1 indicates cooperative binding. Sequencing ladders co-electrophoresed with DNase I footprinting samples were generated with the primers used to amplify the 298-bp fragment.

Fluorescence polarization of ParR–DNA binding. To elucidate the minimal centromere site, ParR binding to the following fluoresceinated oligonucleotides was measured: a 10-bp repeat (AGTATATACT), the 20-bp repeat (AGTATATACTAGTATATACT) and a nonspecific site upstream of the *parC* centromere (TGACATGACATGTAATAAAG). Next, ParR N-terminal domain (ParRN) and ParR C-terminal domain (ParRC) binding to the identified minimal 20-bp site were performed. In all experiments, protein was titrated into a 0.990 ml reaction buffer (25 mM Tris pH 7.5, 100 mM NaCl) containing 1 nM fluorescinated DNA. The resulting data were fitted to a simple bimolecular binding model by nonlinear regression.

Fluorescence polarization experiments of ParM binding to ParR–DNA complexes. We first showed that 1 mM AMP-PCP and 1 mM Mg^{2+} had no effect on the binding of either full-length ParR or ParRN to the fluoresceinated 20-mer (F-20-mer). Next, either full-length ParR-(F-20-mer) or ParRN-(F-20-mer) complexes were formed and ParM was titrated into the reaction buffer (25 mM Tris pH 7.5, 100 mM NaCl, 1 mM AMP-PCP, 1 mM Mg^{2+}) and the data were fitted as before. As a second control, the same experiment was done without full-length ParR or ParRN to check for nonspecific binding of ParM to the F-20-mer DNA site (Supplementary Fig. 6). The slight increase in polarization observed upon addition of high concentrations of ParM into the F-DNA alone was similar to that observed in titrations into preformed ParRN–DNA complex and was therefore due to nonspecific ParM–DNA interactions. When the same experiments were performed without AMP-PCP and Mg^{2+}, no binding was observed to either ParRN–DNA or full-length ParR–DNA.

Electron microscopy. To obtain cryo-electron microscopy images, approximately 3 µL of 20 µM ParR complexed to the 86-bp full-length pSK41 centromere DNA (top strand CAGTATATACTAGTATATACTAAGATATAC-TATAATATATACTGATATATACTAGGATATATCAGTATAGAAAAGTATA-TAAATA) (Fig. 1a) was applied to a glow discharged holey carbon C-flat 1.2/1.3 grid (Protochips) and flash frozen using a Vitrobot (http://fei.com/Vitrobot). Sample grids were transferred to a Gatan 626 cryo-holder and analysed by using a JEOL 2100 cryo-electron microscope operating at 200 kV with a specimen temperature at near liquid N_2 at the National Center for Macromolecular Imaging (Houston, Texas). Electron micrographs were acquired with a Gatan 4k CCD at a magnification of ×60,000 under low-dose conditions.

Structure determination. For crystallization, 20-mer and longer centromeres containing oligonucleotides were used. The best crystals were obtained with a 20-bp site based on a consensus symmetrical site. pSK41 ParR showed no significant sequence homology to any protein and thus molecular replacement could not be used. In addition, ParRN contains no methionine residues, which eliminated the possibility of solving the structure by selenomethionine multi-wavelength anomalous diffraction. Thus, the thymine residues in the DNA site were all singly and multiply substituted with 5-iodouracil. Several substitutions prevented crystallization. Ultimately data were collected for a crystal containing two substitutions (at positions 5 and 7) and a crystal containing four substitutions (at positions 5, 7, 10 and 17). The crystallization solution supplemented with 30% 2-methyl-2,4-petanediol (MPD) served as a cryo-solvent. Because the 20-mer is palindromic, the sites were present on each strand of the duplex. The best data, 3.0 Å resolution, were obtained for the doubly substituted crystal. Data collected on crystals containing no 5-iodouracil substitutions were non-isomorphous with data of the substituted crystals. Thus the twofold substituted crystal was used as 'native'. The two substitutions in the new 'native' were also present as in the fourfold substituted crystal, allowing us to use the latter as a derivative. The locations of the two unique iodo sites in the derivative were located with SOLVE[13]. Density modification produced an electron density map of sufficient quality to build 70% of the model. Phase combination using partial model phases combined with solvent flattening allowed the entire model to be fitted[14]. The final value of R_{free} after refinement was 29.6%[34].

31. Kabsch, W. & Holmes, K. C. The actin fold. *FASEB J.* **9,** 167–174 (1995).
32. Bork, P., Sander, C. & Valencia, A. An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin and hsp70 heat shock proteins. *Proc. Natl Acad. Sci. USA* **89,** 7290–7294 (1992).
33. Kwong, S. M., Skurray, R. A. & Firth, N. *Staphylococcus aureus* multiresistance plasmid pSK41: analysis of the replication region, initiator protein binding and antisense RNA regulation. *Mol. Microbiol.* **51,** 497–509 (2004).
34. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).

ERRATUM

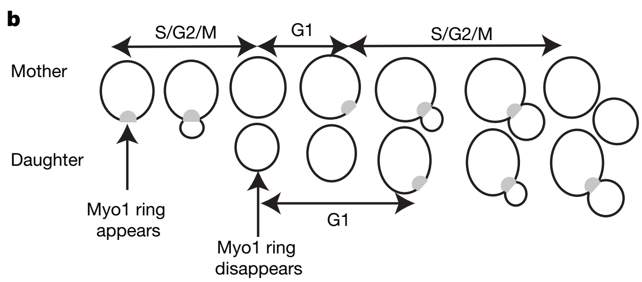
doi:10.1038/nature06511

The effects of molecular noise and size control on variability in the budding yeast cell cycle

Stefano Di Talia, Jan M. Skotheim, James M. Bean, Eric D. Siggia & Frederick R. Cross

Nature 448, 947–951 (2007)

In Fig. 1b of this Letter, the two labels 'S/G2/M' were inadvertently misplaced. The corrected Fig. 1b (in black and white) is shown below.



naturejobs

JOBS OF THE WEEK

Funding is key to the scientific enterprise, but with budgets stretched and ever more proposals being put forward, the way in which the most forward-thinking projects are approved by funding bodies needs to evolve. The US National Institutes of Health (NIH), for example, has been criticized for playing it safe when handing out its grants and, from its own figures, seems more inclined to fund established researchers than new kids on the block. As the agency is the main funder of US biomedical research, this could prevent or discourage young talent, armed with new ideas, from staying in academia.

At the same time, science is evolving and is providing funding agencies with fresh headaches. Many projects now hinge on interdisciplinary and translational research — and multidisciplinary institutes are springing up around the world. Yet there are too few interdisciplinary-minded reviewers in the United States and elsewhere to handle the proposals for this evolving frontier in research.

The NIH is striving to address some of these issues. In the past year it has altered its grant-review system and is pondering further changes. Its 'Director's pioneer' awards and 'Pathway to Independence' programme are aimed at fledgling scientists. Further changes being considered include a shorter time for application processing, reformed study sections and a 'prebuttal' that would let applicants clarify misconceptions about their application before it goes to review (see http://tinyurl.com/38egx4).

All scientists have a stake in debates on how to determine the most promising scientific ideas. The European Union's Framework programmes are famous for an impenetrable grant-application process that requires complex paperwork and encourages, some would say, unwieldy collaborations. Even so, interdisciplinary science is a big part of the most recent Framework's mission (see *Nature* **446,** 104–105; 2007). Science collaborations have become more complex and, in some cases, bureaucracy has become an impediment to funding avant-garde research easily. Grant review needs to evolve along with the scientific enterprise — otherwise that enterprise will not live up to its promise.

Gene Russo, acting editor of *Naturejobs*

CONTACTS
Acting Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
***Nature*events:**
Claudia Paulsen Young
(+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com

France/Switzerland/Belgium:
Muriel Lestringuez (4994)
Southwest UK/RoW:
Nils Moeller (4953)
Scandinavia/Spain/Portugal/Italy:
Evelina Rubio-Hakansson (4973)
Northeast UK/Ireland:
Matthew Ward (+44 (0) 20 7014 4059)
North Germany/The Netherlands:
Reya Silao (4970)
South Germany/Austria:
Hildi Rowland (+44 (0) 20 7014 4084)

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
***Naturejobs* web development:**
Tom Hancock

***Naturejobs* online production:**
Jasmine Myer
US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
Tel: +81-3-3267-8765
e-mail: a.watanabe@natureasia.com

To look too closely

Scope for improvement.

Paul Grainger

The Change Approval Officer, tasked with keeping the computer systems operational, opened the meeting with an edge of annoyance you could shave with. "This extra-ordinary session of Project ERT is to consider the Deep Graphical Department's request for an urgent upgrade to the project's graphics hardware."

"Yes CAO. The Intelligence Department has recently made us aware of an imminent increase in visual surveying requirements."

The CAO sighed with the resignation of a pensioner trying to cross a motorway. "Have you verified the upgrade in the test system?"

"You will understand that we had limited time. This is a reasonably standard upgrade, of which we are generally experienced."

"You fill me with the confidence of a frog in a food-blender. However, the requirement and urgency have been verified. Upgrade approved. Please ensure your back-out plan is more substantial than your resignation."

The two engineers sat in a room filled with monitors, keyboards and Dilbert cartoons. They helped themselves to a jar of pickled chillies as they watched a TV behind their workstations. "Later, a thrilling debate on where we came from and what it's all about between the octogenarian evolution evangelist Richard Dawkins and Pope Albert. But first to NASA's mission control where they're about to bring a multibillion-dollar telescope array online. Other currencies were involved."

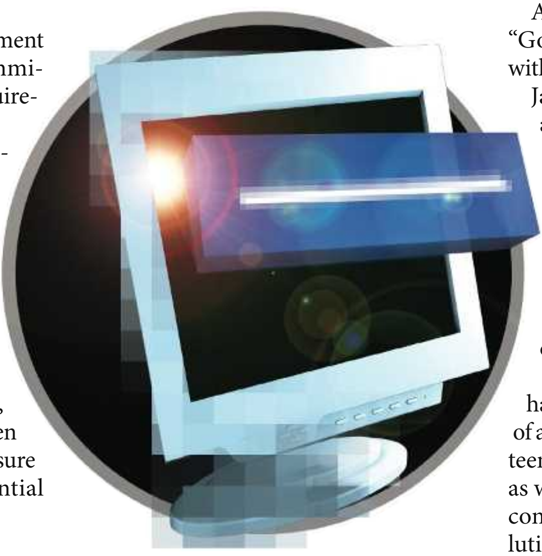
"That's us, Jasper," said one of the engineers.

"This is it, Oddman," said Jasper, sucking the vinegar off a chilli as if it were honey off a bear's paw. "An orbital four-telescope array, huge precision-milled mirrors, grid-based software. We'll be able to see planets, atmospheres, even little aliens." He paused. "Well, maybe not that little."

Oddman swung round to a keyboard and started tapping away. "With what I've done, you'll be able to see up their nostrils, assuming they're standing on their heads, assuming they have heads."

"What have you done?"

Oddman chuckled like a turkey. "I've short cut NASA's CMM level-5 processes, where anything new takes ten years to get in front of a committee, if you're lucky. During our deployment checks, I'll get the five-scope array we've always wanted; I've rewritten the imaging software; we'll be a hundred times better than anyone could have expected." Jasper's mouth dropped open. Oddman put a chilli in it. "Don't worry dude, it's all discrete components, from disparate locations. No one, not even Agent Mulder, could piece this together."



JACEY

The Mission Controller unbuttoned his waistcoat to let the sweat out. "Initial deployment of all four Tycho Brahe telescopes completed. Beginning sequence for configuration checks. Jim, make sure those eggheads at Goddard are ready."

Jim looked up from his Sudoku puzzle. "Just did that Frank. They're watching CNN and eating pickled chillies. They asked if we could send over some herring roll-mops."

"Did you remind them of the procedures they should be working?"

"You bet your head, Frank. I told them NASA never lost a telescope in space and we weren't about to lose one on your watch."

"We never lost a telescope?"

Jim returned to his puzzle. "Don't know Frank. I sort of made that up."

Jasper stared in amazement. "I'll discredit your research if we get caught."

"Chill out, dude. We'll be heroes. They'll name stars after us," replied Oddman, the keyboard now an extension of his hands, like a pint of beer.

Jasper flicked a switch. "Houston, we have go for configuration checks." His voice was as a nervous as a window cleaner who didn't like heights. Or water.

"I'm into the James Webb telescope. Uploading software. I am sex on a stick."

"Realigning array to your new coordinates. The heart of the Universe." Jasper wiped his brow.

"Relax dude, we're about to look farther back in time and with greater detail than anyone else, ever."

Jasper checked his monitor. "Receiving combined signal. This had better work or you're off my Christmas-card list."

A voice crackled over the intercom. "Goddard — we're tracking anomalies with the configuration checks."

Jasper reluctantly flicked his switch again. "We're just looking at that, Houston. We'll get back to you."

Jasper and Oddman looked at each other. Jasper looked as though he'd had a stroke. On the TV behind them, Pope Albert fingered his ring. "God created Heaven and Earth and all creatures upon it. The scripture evidence is substantial."

Dawkins' incredulous stare could have frozen hell, if it existed. "I'm fed up of arguing the toss with such blinkered sixteenth-century retards as you. You might as well believe we're part of an elaborate computer simulation. Evidence for evolution is beyond substantial — it is overwhelming. Your God isn't dead in the face of this evidence — He never existed!"

Pope Albert waved his hand. "He speaks to me. He wants me for a sunbeam."

Oddman broke the stare. "So what can we see?"

Jasper looked at his monitor. It was largely black, with wisps of swirling silver clouds. The clouds suddenly froze. "That's funny. Look in the centre: a small blue rectangle."

"Jesus! What is that?"

"It's growing." Jasper began tapping at his keyboard. "It's growing faster than the speed of light! Your software is as bug-ridden as a camel."

"The software's good."

"Well, if it is, that rectangle will reach us in 30 minutes." Jasper shook his head. "There's some faint white lines across the centre."

"I'll enhance it," Oddman said, doubt beginning to creep over him like malaria.

"It's writing."

Oddman slowly read the writing. "Unrecoverable graphics error. Universe ET™ needs to close down."

"What the hell does that mean?"

"Game over, dude." ■

Paul Grainger is a mid-life crisis waiting to happen. He works in the IT industry to pass the time. He has yet to start a land war in the Middle East.

FUTURES